Gallager's Exponent for MIMO Channels: A Reliability–Rate Tradeoff

Hyundong Shin, *Member, IEEE*, and Moe Z. Win, *Fellow, IEEE*

Corresponding Author:

Hyundong Shin

Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
Room 32-D782, 77 Massachusetts Avenue
Cambridge, MA 02139 USA

Tel.: (617) 253-6173
e-mail: hshin@mit.edu

This research was supported, in part, by the Korea Research Foundation Grant funded by the Korean Government (KRF-2004-214-D00337), the Charles Stark Draper Laboratory Robust Distributed Sensor Networks Program, the Office of Naval Research Young Investigator Award N00014-03-1-0489, and the National Science Foundation under Grant ANI-0335256.

H. Shin and M. Win are with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 USA (e-mail: hshin@mit.edu, moewin@mit.edu).

Abstract

In this paper, we derive Gallager's random coding error exponent for multiple-input multiple-output (MIMO) channels, assuming no channel-state information (CSI) at the transmitter and perfect CSI at the receiver. This measure gives insight into a fundamental tradeoff between the *communication reliability* and *information rate* of MIMO channels, enabling to determine the required codeword length to achieve a prescribed error probability at a given rate below the channel capacity. We quantify the effects of the number of antennas, channel coherence time, and spatial fading correlation on the MIMO exponent. In addition, general formulae for the ergodic capacity and the cutoff rate in the presence of spatial correlation are deduced from the exponent expressions. These formulae are applicable to arbitrary structures of transmit and receive correlation, encompassing all the previously known results as special cases of our expressions.

Index Terms

Block fading, channel capacity, cutoff rate, multiple-input multiple-output (MIMO) system, random coding error exponent, spatial fading correlation.

I. INTRODUCTION

The channel capacity is a crucial information-theoretic perspective that determines the fundamental limit on achievable information rates over a communication channel [1]. However, since the channel capacity alone gives only the knowledge of the maximum achievable rate, a stronger form of the channel coding theorem has been pursued to determine the behavior of the error probability P_{e} as a function of the codeword length N and information rate R [2]–[4]. The *reliability function* or the *error exponent* of a communication system is defined by [2]

$$E\left(R\right) \triangleq \limsup_{N\to\infty} \frac{-\ln P_{\mathrm{e}}^{\mathrm{opt}}\left(R,N\right)}{N}$$

where $P_{\mathrm{e}}^{\mathrm{opt}}\left(R,N\right)$ is the average block error probability for the optimal block code of length N and rate R.[1] The error exponent describes a decaying rate in the error probability as a function of the codeword length, and hence serves to indicate how difficult it may be to achieve a certain level of reliability in communication at a rate below the channel capacity. Although it is difficult to find the exact error exponent, its classical lower bound is available due to Gallager [3]. This

[1]In the following, we will use the term "error probability" to denote the average block error probability.

lower bound is known as the *random coding error exponent* or *Gallager's exponent* in honor of his discovery, and has been used to estimate the codeword length required to achieve a prescribed error probability [5]–[7].

The random coding exponent was extensively studied for single-input single-output (SISO) and single-input multiple-output (SIMO) flat-fading channels with average or peak power constraint [5], [6]. For SIMO block-fading channels, the random coding exponent was derived in [8] with perfect channel-state information (CSI) at the receiver, where it has been shown that although the capacity is independent of the channel coherence time (first asserted in [9] and also recently addressed in [10] and [11] for multiple-antenna communication), the error exponent suffers a considerable decrease due to a reduction in the effective codeword length as the coherence time increases.[2] Therefore, this so-called *channel-incurable effect* reduces the communication reliability. While there are numerous prior investigations (following the seminal work of [12]–[15]) on the capacity for multiple-input multiple-output (MIMO) channels [16]–[23], only limited results are available for error exponents. The random coding exponent were given implicitly in [16] (without final analytical expressions) for independent and identically distributed (i.i.d.) Rayleigh-fading MIMO channels with a single-symbol coherence time, perfect receive CSI, and Gaussian inputs subject to the average power constraint. Also, the random coding exponent was analyzed in [7] for i.i.d. block-fading MIMO channels with no CSI and isotropically unitary inputs subject to the average power constraint.

In this paper, taking into account spatial fading correlation, we derive Gallager's exponent for MIMO channels. We consider a block-fading channel with Gaussian inputs subject to the average power constraint and perfect CSI at the receiver. Our results resort to the methodology developed in [22] and [23], which is based on the finite random matrix theory [24], [25]. The MIMO exponent obtained in the paper provides insight into a fundamental tradeoff between the communication reliability and information rate (below the channel capacity), enabling to determine the required codeword length for a prescribed error probability. It is interesting to note that as a special case of this *reliability–rate* tradeoff, one can obtain the diversity–multiplexing tradeoff of MIMO channels [12], [13], [26], which is a scaled version of the

[2]This observation is parallel to the *divergent* behavior of the channel capacity and cutoff rate of a channel with block memory [9].

asymptotic reliability–rate tradeoff at high signal-to-noise ratio (SNR). We quantify the effects of the number of antennas, the channel coherence time, and the amount of spatial fading correlation on the MIMO exponent. Moreover, the general formulae for the ergodic capacity and cutoff rate are deduced from the exponent expressions. In particular, our capacity formula embraces all the previously known results for i.i.d. [16], [21], one-sided correlated [19], [20], and doubly correlated [22] channels.

This paper is organized as follows. In Section II, signal and channel models are presented. Section III derives the expression for the MIMO random coding exponent. Section IV gives proofs of the main results stated in Theorem 1. In Section V, some numerical results are provided to illustrate the reliability–rate tradeoff in block-fading MIMO channels. Finally, Section VI concludes the paper.

Notation: Throughout the paper, we shall use the following notation. $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ denote the natural numbers and the fields of real and complex numbers, respectively. The superscripts T and † stand for the transpose and transpose conjugate, respectively. $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and (A_{ij}) denotes the matrix with the (i,j)th entry A_{ij}. The trace operator of a square matrix $\boldsymbol{A}$ is denoted by $\mathrm{tr}\left(\boldsymbol{A}\right)$ and $\mathrm{etr}\left(\boldsymbol{A}\right) = e^{\mathrm{tr}(\boldsymbol{A})}$. The Kronecker product of matrices is denoted by $\otimes$. By $\boldsymbol{A} > 0$, we denote $\boldsymbol{A}$ is positive definite. For a Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\lambda_1\left(\boldsymbol{A}\right) \geq \lambda_2\left(\boldsymbol{A}\right) \geq \ldots \geq \lambda_n\left(\boldsymbol{A}\right)$ denotes the eigenvalues of $\boldsymbol{A}$ in decreasing order and $\boldsymbol{\lambda}\left(\boldsymbol{A}\right) \in \mathbb{R}^n$ denote the vector of the ordered eigenvalues, whose ith element is $\lambda_i\left(\boldsymbol{A}\right)$. Also, $\varrho\left(\boldsymbol{A}\right)$ denotes the number of distinct eigenvalues of $\boldsymbol{A}$, and $\lambda_{\langle k\rangle}\left(\boldsymbol{A}\right)$ and $\chi_k\left(\boldsymbol{A}\right)$, $k = 1, 2, \ldots, \varrho\left(\boldsymbol{A}\right)$, denote the distinct eigenvalues of $\boldsymbol{A}$ in decreasing order and its multiplicity, respectively, that is, $\lambda_{\langle 1\rangle}\left(\boldsymbol{A}\right) > \lambda_{\langle 2\rangle}\left(\boldsymbol{A}\right) > \ldots > \lambda_{\langle\varrho(\boldsymbol{A})\rangle}\left(\boldsymbol{A}\right)$ and $\sum_{k=1}^{\varrho(\boldsymbol{A})} \chi_k\left(\boldsymbol{A}\right) = n$. Finally, we shall use the notation $\boldsymbol{X} \in \mathbb{C}^{m\times n} \sim \tilde{\mathcal{N}}_{m,n}\left(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\right)$ to denote that a random matrix $\boldsymbol{X}$ is (matrix-variate) Gaussian distributed with the probability density function (pdf)

$$p_{\boldsymbol{X}}\left(\boldsymbol{X}\right) = \pi^{-mn} \det\left(\boldsymbol{\Sigma}\right)^{-n} \det\left(\boldsymbol{\Psi}\right)^{-m} \mathrm{etr}\left\{-\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}-\boldsymbol{M}\right)\boldsymbol{\Psi}^{-1}\left(\boldsymbol{X}-\boldsymbol{M}\right)^{\dagger}\right\} \tag{1}$$

where $\boldsymbol{M} \in \mathbb{C}^{m\times n}$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\dagger} \in \mathbb{C}^{m\times m} > 0$, and $\boldsymbol{\Psi} = \boldsymbol{\Psi}^{\dagger} \in \mathbb{C}^{n\times n} > 0$.

II. Signal and Channel Models

We consider a MIMO system with n_{T} transmit and n_{R} receive antennas, where the channel remains constant for N_{c} symbol periods and changes independently to a new value for each

coherence time, i.e., every N_{c} symbols. Since the propagation coefficients independently acquire new values for every coherence interval, the channel is memoryless when considering a block length of N_{c} symbols as one channel use with input and output signals of dimension $n_{\mathrm{T}} \times N_{\mathsf{c}}$ and $n_{\mathrm{R}} \times N_{\mathsf{c}}$, respectively.

For an observation interval of $N_{\mathsf{b}} N_{\mathsf{c}}$ symbol periods, the received signal is a sequence $\{\boldsymbol{Y}_k\}_{k=1}^{N_{\mathsf{b}}}$, each $\boldsymbol{Y}_k \in \mathbb{C}^{n_{\mathrm{R}} \times N_{\mathsf{c}}}$ is given by

$$\boldsymbol{Y}_k = \boldsymbol{H}_k \boldsymbol{X}_k + \boldsymbol{W}_k \,, \qquad k = 1, 2, \ldots, N_{\mathsf{b}} \tag{2}$$

where $\boldsymbol{X}_k \in \mathbb{C}^{n_{\mathrm{T}} \times N_{\mathsf{c}}}$ are the transmitted signal matrices, $\boldsymbol{H}_k \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ are the channel matrices, and $\boldsymbol{W}_k \sim \tilde{\mathcal{N}}_{n_{\mathrm{R}}, N_{\mathsf{c}}} \left(\boldsymbol{0}, N_0 \boldsymbol{I}_{n_{\mathrm{R}}}, \boldsymbol{I}_{N_{\mathsf{c}}} \right)$ are the additive white Gaussian noise (AWGN) matrices. Fig. 1 shows a communication link with n_{T} transmit and n_{R} receive antennas to communicate at a rate R (in bits or nats per symbol) over N_{b} independent N_{c}-symbol coherence intervals. Since the channel is memoryless with identical channel statistics for each coherence time interval, the index k can be dropped.

Let $p_{\boldsymbol{X}}\left(\boldsymbol{X}\right)$ be the input probability assignment for $\boldsymbol{X} \in \mathbb{C}^{n_{\mathrm{T}} \times N_{\mathsf{c}}}$ with covariance $\mathbb{C}\mathrm{ov}\left\{\mathrm{vec}\left(\boldsymbol{X}^\dagger\right)\right\} = \boldsymbol{Q}^T \otimes \boldsymbol{I}_{N_{\mathsf{c}}}$ subject to the average power constraint of the form

$$\frac{1}{N_{\mathsf{c}}}\, \mathbb{E}\left\{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\dagger\right)\right\} = \frac{1}{N_{\mathsf{c}}}\, \mathrm{tr}\left(\boldsymbol{Q}^T \otimes \boldsymbol{I}_{N_{\mathsf{c}}}\right) = \mathrm{tr}\left(\boldsymbol{Q}\right) \leq \mathcal{P} \tag{3}$$

where $\boldsymbol{Q}$ is the $n_{\mathrm{T}} \times n_{\mathrm{T}}$ positive semidefinite matrix and $\mathcal{P}$ is the total transmit power over n_{T} transmit antennas. Taking into account spatial fading correlation at both the transmitter and the receiver, we consider the channel matrix $\boldsymbol{H}$ is given by [17], [18]

$$\boldsymbol{H} = \boldsymbol{\Phi}_{\mathrm{R}}^{1/2} \boldsymbol{H}_0 \boldsymbol{\Phi}_{\mathrm{T}}^{1/2} \tag{4}$$

where $\boldsymbol{\Phi}_{\mathrm{T}} \in \mathbb{C}^{n_{\mathrm{T}} \times n_{\mathrm{T}}} > 0$ and $\boldsymbol{\Phi}_{\mathrm{R}} \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{R}}} > 0$ are the transmit and receive correlation matrices, respectively, and $\boldsymbol{H}_0 \sim \tilde{\mathcal{N}}_{n_{\mathrm{R}}, n_{\mathrm{T}}} \left(\boldsymbol{0}, \boldsymbol{I}_{n_{\mathrm{R}}}, \boldsymbol{I}_{n_{\mathrm{T}}} \right)$ is a matrix with i.i.d., zero-mean, unit-variance, complex Gaussian entries. The (i,j) entry H_{ij}, $i = 1, 2, \ldots, n_{\mathrm{R}}$, $j = 1, 2, \ldots, n_{\mathrm{T}}$, of $\boldsymbol{H}$ is a complex propagation coefficient between the jth transmit antenna and the ith receive antenna with $\mathbb{E}\left\{|H_{ij}|^2\right\} = 1$. Note that $\boldsymbol{H} \sim \tilde{\mathcal{N}}_{n_{\mathrm{R}}, n_{\mathrm{T}}} \left(\boldsymbol{0}, \boldsymbol{\Phi}_{\mathrm{R}}, \boldsymbol{\Phi}_{\mathrm{T}} \right)$ [21]. With perfect CSI at the receiver, we have the transition pdf

$$p\left(\boldsymbol{Y} | \boldsymbol{X}, \boldsymbol{H}\right) = \left(\pi N_0\right)^{-n_{\mathrm{R}} N_{\mathsf{c}}} \mathrm{etr}\left\{ -\frac{1}{N_0} \left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\right)\left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\right)^\dagger \right\} \tag{5}$$

which completely characterizes a block-fading MIMO channel.

In what follows, we define the random matrix $\mathbf{\Theta} \in \mathbb{C}^{m\times m} > 0$ as

$$\mathbf{\Theta} \triangleq \begin{cases} \boldsymbol{H}\boldsymbol{H}^{\dagger}, & \text{if } n_{\mathrm{R}} \leq n_{\mathrm{T}} \\ \boldsymbol{H}^{\dagger}\boldsymbol{H}, & \text{otherwise} \end{cases} \tag{6}$$

which is a matrix quadratic form in complex Gaussian matrices, denoted by $\mathbf{\Theta} \sim \tilde{\mathcal{Q}}_{m,n}\left(\boldsymbol{I}_n, \mathbf{\Phi}_1, \mathbf{\Phi}_2\right)$ [21], where $m \triangleq \min\left\{n_{\mathrm{T}}, n_{\mathrm{R}}\right\}$, $n \triangleq \max\left\{n_{\mathrm{T}}, n_{\mathrm{R}}\right\}$, and

$$\left(\mathbf{\Phi}_1 \in \mathbb{C}^{m\times m}, \mathbf{\Phi}_2 \in \mathbb{C}^{n\times n}\right) = \begin{cases} \left(\mathbf{\Phi}_{\mathrm{R}}, \mathbf{\Phi}_{\mathrm{T}}\right), & \text{if } n_{\mathrm{R}} \leq n_{\mathrm{T}} \\ \left(\mathbf{\Phi}_{\mathrm{T}}, \mathbf{\Phi}_{\mathrm{R}}\right), & \text{otherwise.} \end{cases} \tag{7}$$

The pdf of $\mathbf{\Theta} \sim \tilde{\mathcal{Q}}_{m,n}\left(\boldsymbol{I}_n, \mathbf{\Phi}_1, \mathbf{\Phi}_2\right)$ is given by [22]

$$p_{\mathbf{\Theta}}\left(\mathbf{\Theta}\right) = \frac{1}{\tilde{\Gamma}_m\left(n\right)} \det\left(\mathbf{\Phi}_1\right)^{-n} \det\left(\mathbf{\Phi}_2\right)^{-m} \det\left(\mathbf{\Theta}\right)^{n-m} {}_0\tilde{F}_0^{(n)}\left(-\mathbf{\Phi}_1^{-1}\mathbf{\Theta}, \mathbf{\Phi}_2^{-1}\right), \quad \mathbf{\Theta} > 0, \tag{8}$$

where $\tilde{\Gamma}_m\left(\alpha\right) = \pi^{m(m-1)/2}\prod_{i=0}^{m-1}\Gamma\left(\alpha - i\right)$, $\Re\left\{\alpha\right\} > m-1$, is the complex multivariate gamma function, $\Gamma\left(\cdot\right)$ is the Euler gamma function, and ${}_p\tilde{F}_q^{(n)}\left(\cdot\right)$ is the hypergeometric function of two Hermitian matrices, defined by [24, eq. (88)].

III. MIMO Exponent: Reliability–Rate Tradeoff

This section is based on Gallager's random coding bound on the error probability of maximum-likelihood (ML) decoding for a channel with continuous inputs and outputs [3]. Notably, the bound determines the behavior of the error probability as a function of the rate and the codeword length. Hence, by determining Gallager's exponent, we can obtain significant insight into the reliability–rate tradeoff in communication over MIMO channels and the required codeword length to achieve a certain level of reliable communication. In particular, the diversity–multiplexing tradeoff of MIMO channels [12], [13], [26] is a special case of the reliability–rate tradeoff as the SNR goes to infinity.

A. Random Coding Exponent

Using the formulation developed in [3, ch. 7], we obtain the random coding bound on the error probability of ML decoding over block-fading MIMO channels as[3]

$$P_{\mathrm{e}} \leq \left(\frac{2e^{r\delta}}{\xi}\right)^2 e^{-N_{\mathrm{b}}N_{\mathrm{c}}\,E_{\mathrm{r}}(p_{\boldsymbol{X}}(\boldsymbol{X}),R,N_{\mathrm{c}})} \tag{9}$$

[3]When $\boldsymbol{X} = (X_{ij})$ is an $m \times n$ matrix of complex variables that do not depend functionally on each other, $d\boldsymbol{X} = \prod_{i=1}^{m}\prod_{j=1}^{n} d\Re X_{ij}\, d\Im X_{ij}$. If $\boldsymbol{X} \in \mathbb{C}^{m\times m}$ is Hermitian, then $d\boldsymbol{X} = \prod_{i=1}^{m} dX_{ii}\prod_{i<j}^{m} d\Re X_{ij}\, d\Im X_{ij}$.

where $r, \delta \geq 0$ and

$$\xi \approx \frac{\delta}{\sqrt{2\pi N_{\mathrm{b}} \sigma_{\xi}^{2}}} \tag{10}$$

$$\sigma_{\xi}^{2} = \int_{\boldsymbol{X}} \left[\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\dagger}\right) - N_{\mathrm{c}}\mathcal{P}\right]^{2} p_{\boldsymbol{X}}\left(\boldsymbol{X}\right) d\boldsymbol{X}. \tag{11}$$

The random coding exponent $E_{\mathrm{r}}\left(p_{\boldsymbol{X}}\left(\boldsymbol{X}\right), R, N_{\mathrm{c}}\right)$ in (9) is given by

$$E_{\mathrm{r}}\left(p_{\boldsymbol{X}}\left(\boldsymbol{X}\right), R, N_{\mathrm{c}}\right) = \max_{0 \leq \rho \leq 1} \left\{ \max_{r \geq 0} E_{0}\left(p_{\boldsymbol{X}}\left(\boldsymbol{X}\right), \rho, r, N_{\mathrm{c}}\right) - \rho R \right\} \tag{12}$$

with

$$\begin{aligned} & E_{0}\left(p_{\boldsymbol{X}}\left(\boldsymbol{X}\right), \rho, r, N_{\mathrm{c}}\right) \\ & = -\frac{1}{N_{\mathrm{c}}} \ln \left\{ \int_{\boldsymbol{H}} p_{\boldsymbol{H}}\left(\boldsymbol{H}\right) \int_{\boldsymbol{Y}} \left(\int_{\boldsymbol{X}} p_{\boldsymbol{X}}\left(\boldsymbol{X}\right) e^{r\left[\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\dagger}\right) - N_{\mathrm{c}}\mathcal{P}\right]} p\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{H}\right)^{1/(1+\rho)} d\boldsymbol{X} \right)^{1+\rho} d\boldsymbol{Y}\, d\boldsymbol{H} \right\}. \end{aligned} \tag{13}$$

The parameter r to be optimized may be viewed as a Lagrange multiplier corresponding to the input-power constraint [7].

1) Capacity-Achieving Input Distribution: As in [3]–[8], we choose the capacity-achieving distribution for $p_{\boldsymbol{X}}\left(\boldsymbol{X}\right)$ satisfying the power constraint (3), namely,

$$p_{\boldsymbol{X}}\left(\boldsymbol{X}\right) = \pi^{-n_{\mathrm{T}} N_{\mathrm{c}}} \det\left(\boldsymbol{Q}\right)^{-N_{\mathrm{c}}} \mathrm{etr}\left(-\boldsymbol{Q}^{-1}\boldsymbol{X}\boldsymbol{X}^{\dagger}\right) \tag{14}$$

with $\mathrm{tr}\left(\boldsymbol{Q}\right) \leq \mathcal{P}$.[4] Although this choice of the Gaussian input distribution is optimal only if the rate approaches the channel capacity, it makes the problem analytically tractable [3].

Proposition 1: Let $E_{0,\tilde{\mathcal{N}}}\left(\boldsymbol{Q}, \rho, r, N_{\mathrm{c}}\right)$ be $E_{0}\left(p_{\boldsymbol{X}}\left(\boldsymbol{X}\right), \rho, r, N_{\mathrm{c}}\right)$ in (13) for the Gaussian input distribution $p_{\boldsymbol{X}}\left(\boldsymbol{X}\right)$ of (14). Then, we have

$$\begin{aligned} E_{0,\tilde{\mathcal{N}}}\left(\boldsymbol{Q}, \rho, r, N_{\mathrm{c}}\right) &= r\mathcal{P}\left(1+\rho\right) + \left(1+\rho\right) \ln\det\left(\boldsymbol{I}_{n_{\mathrm{T}}} - r\boldsymbol{Q}\right) \\ & \quad - \frac{1}{N_{\mathrm{c}}} \ln \mathbb{E}\left\{ \det\left(\boldsymbol{I}_{n_{\mathrm{R}}} + \frac{\boldsymbol{H}\left(\boldsymbol{Q}^{-1} - r\boldsymbol{I}_{n_{\mathrm{T}}}\right)^{-1}\boldsymbol{H}^{\dagger}}{N_{0}\left(1+\rho\right)} \right)^{-N_{\mathrm{c}}\rho} \right\}. \end{aligned} \tag{15}$$

Proof: See Appendix A. □

[4]In general, optimization of the input distribution $p_{\boldsymbol{X}}\left(\boldsymbol{X}\right)$ under the power constraint (3) to maximize the error exponent (i.e., to minimize the upper bound) is a difficult task.

For the case of equal power allocation to each of transmit antennas, i.e., $\boldsymbol{Q} = \frac{\mathcal{P}}{n_{\mathrm{T}}}\boldsymbol{I}_{n_{\mathrm{T}}}$ (because the transmitter has no channel knowledge), (15) becomes

$$E_{0,\tilde{\mathcal{N}}}\left(\tfrac{\mathcal{P}}{n_{\mathrm{T}}}\boldsymbol{I}_{n_{\mathrm{T}}}, \rho, r, N_{\mathrm{c}}\right) = r\mathcal{P}\left(1+\rho\right) + n_{\mathrm{T}}\left(1+\rho\right)\ln\left(\frac{n_{\mathrm{T}} - r\mathcal{P}}{n_{\mathrm{T}}}\right) - \frac{1}{N_{\mathrm{c}}}\ln\mathbb{E}\left\{\det\left(\boldsymbol{I}_{n_{\mathrm{R}}} + \frac{\gamma \boldsymbol{H}\boldsymbol{H}^{\dagger}}{\left(n_{\mathrm{T}} - r\mathcal{P}\right)\left(1+\rho\right)}\right)^{-N_{\mathrm{c}}\rho}\right\} \tag{16}$$

where $\gamma = \mathcal{P}/N_0$ is the average SNR at each receive antenna. Let us introduce a new variable $\beta = n_{\mathrm{T}} - r\mathcal{P}$ where β is restricted to the range $0 \le \beta \le n_{\mathrm{T}}$ to have a meaningful result in (16). Then, we have

$$\begin{aligned}\tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right) &\triangleq E_{0,\tilde{\mathcal{N}}}\left(\tfrac{\mathcal{P}}{n_{\mathrm{T}}}\boldsymbol{I}_{n_{\mathrm{T}}}, \rho, r, N_{\mathrm{c}}\right)\Big|_{\beta = n_{\mathrm{T}} - r\mathcal{P}} \\ &= \underbrace{\left(1+\rho\right)\left(n_{\mathrm{T}} - \beta\right) + n_{\mathrm{T}}\left(1+\rho\right)\ln\left(\beta/n_{\mathrm{T}}\right)}_{\triangleq \mathcal{K}(\rho,\beta)} - \frac{1}{N_{\mathrm{c}}}\ln\mathcal{L}_0\left(\rho, \beta, N_{\mathrm{c}}\right)\end{aligned} \tag{17}$$

where

$$\mathcal{L}_0\left(\rho, \beta, N_{\mathrm{c}}\right) = \mathbb{E}\left\{\det\left(\boldsymbol{I}_m + \frac{\gamma\boldsymbol{\Theta}}{\beta\left(1+\rho\right)}\right)^{-N_{\mathrm{c}}\rho}\right\}. \tag{18}$$

With maximization over $\beta \in [0, n_{\mathrm{T}}]$ and $\rho \in [0, 1]$ to obtain the tightest bound, we have the random coding exponent for Gaussian codebooks and equal power allocation as follows:[5]

$$\begin{aligned}E_{\mathrm{r}}\left(R, N_{\mathrm{c}}\right) &\triangleq E_{\mathrm{r}}\left(p_{\boldsymbol{X}}\left(\boldsymbol{X}\right), R, N_{\mathrm{c}}\right)\Big|_{\boldsymbol{X}\sim\tilde{\mathcal{N}}_{n_{\mathrm{T}}, N_{\mathrm{c}}}\left(\boldsymbol{0}, \frac{\mathcal{P}}{n_{\mathrm{T}}}\boldsymbol{I}_{n_{\mathrm{T}}}, \boldsymbol{I}_{N_{\mathrm{c}}}\right)} \\ &= \max_{0\le\rho\le1}\left\{\max_{0\le\beta\le n_{\mathrm{T}}}\tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right) - \rho R\right\}.\end{aligned} \tag{19}$$

Proposition 2: Let $\beta^*\left(\rho\right)$ be the value of β that maximizes $\tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right)$ defined in (17) for each $\rho \in [0, 1]$. Then, $\beta^*\left(\rho\right)$ is the solution of $\partial\tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right)/\partial\beta = 0$ and is always in the range $0 < \beta \le n_{\mathrm{T}}$.

Proof: See Appendix B. □

It can be shown using (64) and (67) in Appendix B that as $\gamma \to \infty$ or $\gamma \to 0$, the optimal value of β does not depend on N_{c}, that is,

$$\lim_{\gamma\to\infty}\beta^*\left(\rho\right) = n_{\mathrm{T}} - \frac{m\rho}{1+\rho} \quad \text{and} \quad \lim_{\gamma\to 0}\beta^*\left(\rho\right) = n_{\mathrm{T}}.$$

[5]The random coding bound can be improved by expurgating "bad" codewords from the code ensemble at low rates (see, e.g, [3]). More details for the expurgated exponent of block-fading MIMO channels can be found in [27].

According to Proposition 2 and using the general relation $dE_{\mathrm{r}}\left(R, N_{\mathrm{c}}\right)/dR = -\rho$, the maximization of the exponent in (19) over $\beta \in [0, n_{\mathrm{T}}]$ and $\rho \in [0,1]$ can be performed by the following parametric equations:

$$E_{\mathrm{r}}\left(R, N_{\mathrm{c}}\right) = \tilde{E}_0\left(\rho, \beta^*\left(\rho\right), N_{\mathrm{c}}\right) - \rho R \tag{20}$$

$$R = \left[\frac{\partial \tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right)}{\partial \rho}\right]\Bigg|_{\beta=\beta^*(\rho)} \tag{21}$$

with

$$\frac{\partial \tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right)}{\partial \rho} = \underbrace{\left(n_{\mathrm{T}} - \beta\right) + n_{\mathrm{T}} \ln\left(\beta/n_{\mathrm{T}}\right)}_{\triangleq\, \mathcal{K}^{(\rho)}(\rho,\beta) = \frac{\partial \mathcal{K}(\rho,\beta)}{\partial \rho}} - \frac{1}{N_{\mathrm{c}}} \mathcal{L}_0^{-1}\left(\rho, \beta, N_{\mathrm{c}}\right) \frac{\partial \mathcal{L}_0\left(\rho, \beta, N_{\mathrm{c}}\right)}{\partial \rho} \tag{22}$$

where

$$\frac{\partial \mathcal{L}_0\left(\rho, \beta, N_{\mathrm{c}}\right)}{\partial \rho} = \mathbb{E}\left\{ N_{\mathrm{c}} \det\left(\tfrac{1}{\beta}\boldsymbol{\Omega}_{\rho,\beta}\right)^{-N_{\mathrm{c}}\rho} \left[\tfrac{\rho\gamma}{\beta(1+\rho)^2} \operatorname{tr}\left\{\boldsymbol{\Theta}\left(\tfrac{1}{\beta}\boldsymbol{\Omega}_{\rho,\beta}\right)^{-1}\right\} - \ln\det\left(\tfrac{1}{\beta}\boldsymbol{\Omega}_{\rho,\beta}\right)\right]\right\}. \tag{23}$$

2) Key Quantities: The values of R in (21) at $\rho = 1$ and $\rho = 0$ are the ***critical rate*** R_{cr} and the ***ergodic capacity*** $\langle C \rangle$ of the channel, respectively [3]–[6]. From $\partial \tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right)/\partial \beta$ in (67), we see that $\beta^*\left(0\right) = n_{\mathrm{T}}$ and hence, the ergodic capacity can be written as

$$\langle C \rangle = \left[\frac{\partial \tilde{E}_0\left(\rho, \beta, N_{\mathrm{c}}\right)}{\partial \rho}\right]\Bigg|_{\rho=0,\, \beta=n_{\mathrm{T}}} \tag{24}$$

$$= \mathbb{E}\left\{\ln\det\left(\boldsymbol{I}_m + \tfrac{\gamma}{n_{\mathrm{T}}}\boldsymbol{\Theta}\right)\right\}. \tag{25}$$

We remark that the capacity expression (25) obtained from the exponent is independent of the channel coherence time N_{c} and is in agreement with the previous result [14]–[16]. Also, the quantity E_0 is defined as the value of the exponent $E_{\mathrm{r}}\left(R, N_{\mathrm{c}}\right)$ at $R = 0$, referred to as the ***exponential error-bound parameter*** [4], [5], and is given by $\tilde{E}_0\left(1, \beta^*\left(1\right), N_{\mathrm{c}}\right)$. This quantity is equal to the value of R at which the exponent becomes zero by setting $\rho = 1$ and $\beta = \beta^*\left(1\right)$. If setting $r = 0$ or equivalently $\beta = n_{\mathrm{T}}$ (i.e., without the constraint on the minimum energy of the codewords) in (13), E_0 becomes equal to the ***cutoff rate*** R_0 of the channel

$$R_0 = \tilde{E}_0\left(1, n_{\mathrm{T}}, N_{\mathrm{c}}\right) \tag{26}$$

$$= -\frac{1}{N_{\mathrm{c}}} \ln \mathbb{E}\left\{\det\left(\boldsymbol{I}_m + \tfrac{\gamma}{2n_{\mathrm{T}}}\boldsymbol{\Theta}\right)^{-N_{\mathrm{c}}}\right\}. \tag{27}$$

This is an important parameter, as it determines both the magnitude of the zero-rate exponent and the rate regime in which the error probability can be made arbitrarily small by increasing the codeword length.

3) Effect of Channel Coherence—Channel-Incurable Effect: Using Jensen's inequality, it is easy to show

$$\frac{1}{N_{\mathrm{c}}}\ln\mathcal{L}_0\left(\rho,\beta,N_{\mathrm{c}}\right)\geq\frac{1}{N_{\mathrm{c}}-1}\ln\mathcal{L}_0\left(\rho,\beta,N_{\mathrm{c}}-1\right) \tag{28}$$

yielding

$$\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)\leq\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}-1\right). \tag{29}$$

Therefore, for fixed R, the random coding exponent decreases with N_{c}, while the channel capacity is independent of N_{c}. This reliability reduction is due to the fact that the increase in N_{c} results in a decrease in the number of independent channel realizations across the code and hence, reduces the effectiveness of channel coding to mitigate unfavorable fading. We call this effect of the channel coherence time on communication reliability "a channel-incurable effect". In particular, since $\lim_{N_{\mathrm{c}}\to\infty}\frac{1}{N_{\mathrm{c}}}\ln\mathcal{L}_0\left(\rho,\beta,N_{\mathrm{c}}\right)=0$, we have

$$\lim_{N_{\mathrm{c}}\to\infty}\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)=\mathcal{K}\left(\rho,\beta\right) \tag{30}$$

leading to $\lim_{N_{\mathrm{c}}\to\infty}\beta^*\left(\rho\right)=n_{\mathrm{T}}$ and $\lim_{N_{\mathrm{c}}\to\infty}E_{\mathrm{r}}\left(R,N_{\mathrm{c}}\right)=0$. Therefore, if $N_{\mathrm{c}}\to\infty$, it is impossible to transmit information at any positive rate with arbitrary reliability even with the use of multiple antennas. In fact, n_{T} must also increase without limit so that the so-called *space–time autocoding effect* takes place, which makes arbitrarily reliable communications possible [11].

B. Evaluation of $\tilde{E}_0\left(\rho,\beta,N_c\right)$, $\partial\tilde{E}_0\left(\rho,\beta,N_c\right)/\partial\beta$, and $\partial\tilde{E}_0\left(\rho,\beta,N_c\right)/\partial\rho$

To calculate the random coding exponent, the quantities $\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)$, $\partial\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)/\partial\beta$, and $\partial\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)/\partial\rho$ need to be determined. We now evaluate them in the following theorem which will be proven in the next section.

Theorem 1: Let $\boldsymbol{H}\sim\tilde{\mathcal{N}}_{n_{\mathrm{R}},n_{\mathrm{T}}}\left(\mathbf{0},\boldsymbol{\Phi}_{\mathrm{R}},\boldsymbol{\Phi}_{\mathrm{T}}\right)$ or $\boldsymbol{\Theta}\sim\tilde{\mathcal{Q}}_{m,n}\left(\boldsymbol{I}_n,\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2\right)$. Then,

1) $\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)$ is given by

$$\tilde{E}_0\left(\rho,\beta,N_{\mathrm{c}}\right)=\begin{cases}\mathcal{K}\left(\rho,\beta\right)-\frac{1}{N_{\mathrm{c}}}\ln\left(K_{\mathrm{cor}}^{-1}\det\begin{bmatrix}\mathsf{G}_{(m-N_{\mathrm{c}}\rho)}\left(\boldsymbol{\Phi}_1\right)\\\boldsymbol{\Xi}\left(\rho,\beta\right)\end{bmatrix}\right), & \text{if } N_{\mathrm{c}}\rho\in\{1,2,\ldots,m\}\\ \mathcal{K}\left(\rho,\beta\right)+\frac{\mathcal{T}_{\mathrm{A}}}{N_{\mathrm{c}}}\ln\left(\frac{\gamma}{\beta(1+\rho)}\right) & \\ \qquad-\frac{1}{N_{\mathrm{c}}}\ln\left(\mathcal{T}_{\mathrm{B}}\left(\rho,N_{\mathrm{c}}\right)\det\begin{bmatrix}\mathsf{G}_{(n-m)}\left(\boldsymbol{\Phi}_2\right)\\\boldsymbol{\Upsilon}\left(\rho,\beta\right)\end{bmatrix}\right), & \text{otherwise.}\end{cases} \tag{31}$$

If $\mathbf{\Phi}_{\mathrm{T}} = \boldsymbol{I}_{n_{\mathrm{T}}}$ and $\mathbf{\Phi}_{\mathrm{R}} = \boldsymbol{I}_{n_{\mathrm{R}}}$ (i.i.d. MIMO channel), then $\tilde{E}_0(\rho, \beta, N_{\mathsf{c}})$ reduces to

$$\tilde{E}_0(\rho, \beta, N_{\mathsf{c}}) = \mathcal{K}(\rho, \beta) - \frac{1}{N_{\mathsf{c}}} \ln\Big(K_{\text{iid}}^{-1} \det \mathbf{\Upsilon}_{\text{iid}}(\rho, \beta)\Big). \tag{32}$$

2) $\partial \tilde{E}_0(\rho, \beta, N_{\mathsf{c}}) / \partial \beta$ is given by

$$\frac{\partial \tilde{E}_0(\rho, \beta, N_{\mathsf{c}})}{\partial \beta} = \mathcal{K}^{(\beta)}(\rho, \beta) - \frac{\mathcal{T}_{\mathrm{A}}}{N_{\mathsf{c}}\beta} - \frac{1}{N_{\mathsf{c}}} \operatorname{tr}\left\{\begin{bmatrix} \mathsf{G}_{(n-m)}(\mathbf{\Phi}_2) \\ \mathbf{\Upsilon}(\rho, \beta) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{\Upsilon}^{(\beta)}(\rho, \beta) \end{bmatrix}\right\}. \tag{33}$$

If $\mathbf{\Phi}_{\mathrm{T}} = \boldsymbol{I}_{n_{\mathrm{T}}}$ and $\mathbf{\Phi}_{\mathrm{R}} = \boldsymbol{I}_{n_{\mathrm{R}}}$, then

$$\frac{\partial \tilde{E}_0(\rho, \beta, N_{\mathsf{c}})}{\partial \beta} = \mathcal{K}^{(\beta)}(\rho, \beta) - \frac{1}{N_{\mathsf{c}}} \operatorname{tr}\left\{\mathbf{\Upsilon}_{\text{iid}}^{-1}(\rho, \beta)\, \mathbf{\Upsilon}_{\text{iid}}^{(\beta)}(\rho, \beta)\right\}. \tag{34}$$

3) $\partial \tilde{E}_0(\rho, \beta, N_{\mathsf{c}}) / \partial \rho$ is given by

$$\begin{aligned} \frac{\partial \tilde{E}_0(\rho, \beta, N_{\mathsf{c}})}{\partial \rho} &= \mathcal{K}^{(\rho)}(\rho, \beta) - \frac{\mathcal{T}_{\mathrm{A}}}{N_{\mathsf{c}}(1+\rho)} - \sum_{i=1}^{\varrho(\mathbf{\Phi}_1)} \sum_{j=1}^{\chi_i(\mathbf{\Phi}_1)-1} \frac{j}{N_{\mathsf{c}}\rho - m + \chi_i(\mathbf{\Phi}_1) - j} \\ &\quad + \sum_{k=1}^{m-1} \frac{k}{N_{\mathsf{c}}\rho - k} - \frac{1}{N_{\mathsf{c}}} \operatorname{tr}\left\{\begin{bmatrix} \mathsf{G}_{(n-m)}(\mathbf{\Phi}_2) \\ \mathbf{\Upsilon}(\rho, \beta) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{\Upsilon}^{(\rho)}(\rho, \beta) \end{bmatrix}\right\}, \\ &\qquad N_{\mathsf{c}}\rho \neq 1, 2, \ldots, m-1. \end{aligned} \tag{35}$$

If $\mathbf{\Phi}_{\mathrm{T}} = \boldsymbol{I}_{n_{\mathrm{T}}}$ and $\mathbf{\Phi}_{\mathrm{R}} = \boldsymbol{I}_{n_{\mathrm{R}}}$, then

$$\frac{\partial \tilde{E}_0(\rho, \beta, N_{\mathsf{c}})}{\partial \rho} = \mathcal{K}^{(\rho)}(\rho, \beta) - \frac{1}{N_{\mathsf{c}}} \operatorname{tr}\left\{\mathbf{\Upsilon}_{\text{iid}}^{-1}(\rho, \beta)\, \mathbf{\Upsilon}_{\text{iid}}^{(\rho)}(\rho, \beta)\right\}. \tag{36}$$

The quantities K_{cor}, K_{iid}, $\mathcal{T}_{\mathrm{A}}$, $\mathcal{T}_{\mathrm{B}}(\rho, N_{\mathsf{c}})$, and the matrices $\mathsf{G}_{(\cdot)}(\cdot)$, $\mathbf{\Xi}(\rho, \beta)$, $\mathbf{\Upsilon}(\rho, \beta)$, $\mathbf{\Upsilon}^{(\beta)}(\rho, \beta)$, $\mathbf{\Upsilon}^{(\rho)}(\rho, \beta)$, $\mathbf{\Upsilon}_{\text{iid}}(\rho, \beta)$, $\mathbf{\Upsilon}_{\text{iid}}^{(\beta)}(\rho, \beta)$, and $\mathbf{\Upsilon}_{\text{iid}}^{(\rho)}(\rho, \beta)$ are given in Table I.

Corollary 1 (Ergodic Capacity): If $\boldsymbol{H} \sim \tilde{\mathcal{N}}_{n_{\mathrm{R}}, n_{\mathrm{T}}}(\mathbf{0}, \mathbf{\Phi}_{\mathrm{R}}, \mathbf{\Phi}_{\mathrm{T}})$, then the ergodic capacity $\langle C \rangle$ is given by

$$\langle C \rangle = \operatorname{tr}\left\{\begin{bmatrix} \mathsf{G}_{(n-m)}(\mathbf{\Phi}_2) \\ \mathbf{\Upsilon}(0, n_{\mathrm{T}}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda} \end{bmatrix}\right\} - (m-1) + \sum_{i=1}^{\varrho(\mathbf{\Phi}_1)} \sum_{j=1}^{\chi_i(\mathbf{\Phi}_1)-1} \frac{j}{m - \chi_i(\mathbf{\Phi}_1) + j} \tag{37}$$

with $\mathbf{\Lambda} \in \mathbb{R}^{m \times n}$ given by

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{1,1} & \cdots & \mathbf{\Lambda}_{1,\varrho(\mathbf{\Phi}_2)} \\ \vdots & \ddots & \vdots \\ \mathbf{\Lambda}_{\varrho(\mathbf{\Phi}_1),1} & \cdots & \mathbf{\Lambda}_{\varrho(\mathbf{\Phi}_1),\varrho(\mathbf{\Phi}_2)} \end{bmatrix} \tag{38}$$

where the (i,j)th entry $\Lambda_{p,q,ij}$ of $\mathbf{\Lambda}_{p,q} \in \mathbb{R}^{\chi_p(\mathbf{\Phi}_1)\times\chi_q(\mathbf{\Phi}_2)}$, $p=1,\ldots,\varrho(\mathbf{\Phi}_1)$, $q=1,\ldots,\varrho(\mathbf{\Phi}_2)$, is

$$\Lambda_{p,q,ij} = \mathcal{G}_{i+j-1,2}\left(\tfrac{\gamma}{n_\mathrm{T}}\,\lambda_{\langle p\rangle}(\mathbf{\Phi}_1), \lambda_{\langle q\rangle}(\mathbf{\Phi}_2), m-i+1\right). \tag{39}$$

Proof: Note that

$$\begin{aligned}\mathrm{tr}\left\{\begin{bmatrix}\mathsf{G}_{(n-m)}(\mathbf{\Phi}_2)\\ \mathbf{\Upsilon}(0,n_\mathrm{T})\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\ \mathbf{\Upsilon}^{(\rho)}(0,n_\mathrm{T})\end{bmatrix}\right\} &= -N_\mathsf{c}\,\mathrm{tr}\left\{\begin{bmatrix}\mathsf{G}_{(n-m)}(\mathbf{\Phi}_2)\\ \mathbf{\Upsilon}(0,n_\mathrm{T})\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\ \mathbf{\Lambda}\end{bmatrix}\right\}\\ &\quad+\underbrace{n_\mathrm{T}\,\mathrm{tr}\left\{\begin{bmatrix}\mathsf{G}_{(n-m)}(\mathbf{\Phi}_2)\\ \mathbf{\Upsilon}(0,n_\mathrm{T})\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\ \mathbf{\Upsilon}^{(\beta)}(0,n_\mathrm{T})\end{bmatrix}\right\}}_{=-\mathcal{T}_\mathrm{A}}.\end{aligned} \tag{40}$$

The proof follows immediately from (24), Theorem 1.3 with $\rho=0$ and $\beta=n_\mathrm{T}$, and (40). □

Note that the expression (37) for the ergodic capacity $\langle C\rangle$ is sufficiently general and applicable to arbitrary structures of correlation matrices $\mathbf{\Phi}_\mathrm{T}$ and $\mathbf{\Phi}_\mathrm{R}$, and hence, embraces all the previously known results for i.i.d. channels ($\mathbf{\Phi}_\mathrm{T}=\boldsymbol{I}_{n_\mathrm{T}}$, $\mathbf{\Phi}_\mathrm{R}=\boldsymbol{I}_{n_\mathrm{R}}$) [16], [21], one-sided correlated channels ($\mathbf{\Phi}_1=\boldsymbol{I}_m$ [19] or $\mathbf{\Phi}_2=\boldsymbol{I}_n$ [20]), and doubly correlated channels [22] (where all the eigenvalues of $\mathbf{\Phi}_\mathrm{T}$ and $\mathbf{\Phi}_\mathrm{R}$ are assumed to be distinct) as special cases of (37).

Corollary 2 (Cutoff Rate): If $\boldsymbol{H}\sim\tilde{\mathcal{N}}_{n_\mathrm{R},n_\mathrm{T}}(\mathbf{0},\mathbf{\Phi}_\mathrm{R},\mathbf{\Phi}_\mathrm{T})$, then the cutoff rate R_0 is given by

$$R_0 = \begin{cases} -\frac{1}{N_\mathsf{c}}\ln\left(K_\mathsf{cor}^{-1}\,\det\begin{bmatrix}\mathsf{G}_{(m-N_\mathsf{c})}(\mathbf{\Phi}_1)\\ \mathbf{\Xi}(1,n_\mathrm{T})\end{bmatrix}\right), & \text{if } N_\mathsf{c}\in\{1,2,\ldots,m\}\\[2ex] \frac{\mathcal{T}_\mathrm{A}}{N_\mathsf{c}}\ln\left(\frac{\gamma}{2n_\mathrm{T}}\right)-\frac{1}{N_\mathsf{c}}\ln\left(\mathcal{T}_\mathrm{B}(1,N_\mathsf{c})\det\begin{bmatrix}\mathsf{G}_{(n-m)}(\mathbf{\Phi}_2)\\ \mathbf{\Upsilon}(1,n_\mathrm{T})\end{bmatrix}\right), & \text{otherwise.}\end{cases} \tag{41}$$

In particular, if $\mathbf{\Phi}_\mathrm{T}=\boldsymbol{I}_{n_\mathrm{T}}$ and $\mathbf{\Phi}_\mathrm{R}=\boldsymbol{I}_{n_\mathrm{R}}$, then we have

$$R_0 = -\frac{1}{N_\mathsf{c}}\ln\Big(K_\mathsf{iid}^{-1}\,\det\mathbf{\Upsilon}_\mathsf{iid}(1,n_\mathrm{T})\Big). \tag{42}$$

Proof: It follows immediately from (26) and Theorem 1.1 with $\rho=1$ and $\beta=n_\mathrm{T}$. □

C. Coding Requirement

As in [6], we can approximate the required codeword length to achieve a prescribed error probability P_e at a rate R by solving for N_b in the following equation:

$$P_\mathsf{e} = \left(2e^{r\delta}/\xi\right)^2 e^{-N_\mathsf{b}N_\mathsf{c}E_\mathsf{r}(R,N_\mathsf{c})}. \tag{43}$$

Using (10), it is easy to see that the factor $\left(2e^{r\delta}/\xi\right)^2$ in (43) is minimized over $\delta \geq 0$, for large N_b, by choosing $\delta = 1/r$ [3]. This yields

$$\min_{\delta \geq 0} \left(2e^{r\delta}/\xi\right)^2 \approx 8\pi e^2 \sigma_\xi^2 r^2 N_\mathrm{b} \quad \text{for large } N_\mathrm{b}. \tag{44}$$

Also, from (11) and [23, Lemma 5], we have

$$\sigma_\xi^2 = N_\mathrm{c} \mathcal{P}^2 / n_\mathrm{T}. \tag{45}$$

Combining (44) and (45) together with the fact that $\beta = n_\mathrm{T} - r\mathcal{P}$, (43) can be written as

$$P_\mathrm{e} = (8\pi/n_\mathrm{T}) \left\{n_\mathrm{T} - \beta^*(\rho)\right\}^2 N_\mathrm{b} N_\mathrm{c}\, e^{-N_\mathrm{b} N_\mathrm{c} E_\mathrm{r}(R, N_\mathrm{c}) + 2}. \tag{46}$$

After solving for N_b in (46), we take $L = N_\mathrm{c} \cdot \lceil N_\mathrm{b} \rceil$ as our estimate of the codeword length (in symbol) required to achieve P_e at the rate R, where $\lceil \cdot \rceil$ denotes the smallest integer larger than or equal an enclosed quantity.

IV. Proof of the Main Theorem

In this section, we provide proofs of the main results stated in Theorem 1. The methodology recently developed in [22] and [23] for dealing with random matrices paves a way to prove the theorem.

A. Proof of Theorem 1.1

Using the same steps leading to [22, Theorem 1], we get

$$\begin{aligned}
\mathcal{L}_0(\rho, \beta, N_\mathrm{c}) &= \int_{\boldsymbol{\Theta} = \boldsymbol{\Theta}^\dagger > 0} \det\left(\boldsymbol{I}_m + \eta \boldsymbol{\Theta}\right)^{-N_\mathrm{c}\rho} p_{\boldsymbol{\Theta}}(\boldsymbol{\Theta})\, d\boldsymbol{\Theta} \\
&= \frac{\pi^{m(m-1)} \det\left(\boldsymbol{\Phi}_2\right)^{-m}}{\tilde{\Gamma}_m(n)\, \tilde{\Gamma}_m(m)} \int_{\boldsymbol{\lambda}(\boldsymbol{\Theta})} \prod_{k=1}^{m} \lambda_k^{n-m}(\boldsymbol{\Theta}) \prod_{i<j}^{m} \left(\lambda_i(\boldsymbol{\Theta}) - \lambda_j(\boldsymbol{\Theta})\right)^2 \\
&\qquad \times {}_1\tilde{F}_0^{(m)}\left(N_\mathrm{c}\rho; \boldsymbol{D}, -\eta\, \boldsymbol{\Phi}_1\right) {}_0\tilde{F}_0^{(n)}\left(\boldsymbol{D}, -\boldsymbol{\Phi}_2^{-1}\right) d\boldsymbol{\lambda}(\boldsymbol{\Theta})
\end{aligned} \tag{47}$$

where $\eta = \frac{\gamma}{\beta(1+\rho)}$ and $\boldsymbol{D} = \operatorname{diag}\left(\lambda_1(\boldsymbol{\Theta}), \lambda_2(\boldsymbol{\Theta}), \ldots, \lambda_m(\boldsymbol{\Theta})\right)$.[6] Successively applying the generic determinantal formula for hypergeometric functions of matrix arguments [23, Lemma 4] and the

[6]For $\boldsymbol{A} = \boldsymbol{A}^\dagger \in \mathbb{C}^{p \times p} > 0$, we denote $\int_{\boldsymbol{\lambda}(\boldsymbol{A})} d\boldsymbol{\lambda}(\boldsymbol{A}) = \int_0^{\lambda_{p-1}(\boldsymbol{A})} \int_{\lambda_p(\boldsymbol{A})}^{\lambda_{p-2}(\boldsymbol{A})} \cdots \int_{\lambda_2(\boldsymbol{A})}^{\infty} d\lambda_1(\boldsymbol{A}) d\lambda_2(\boldsymbol{A}) \cdots d\lambda_p(\boldsymbol{A})$. If the integrand is symmetric in $\lambda_1(\boldsymbol{A}), \lambda_2(\boldsymbol{A}), \ldots, \lambda_p(\boldsymbol{A})$, then

$$\int_{\boldsymbol{\lambda}(\boldsymbol{A})} d\boldsymbol{\lambda}(\boldsymbol{A}) = \frac{1}{p!} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{p\text{-fold}} d\lambda_1(\boldsymbol{A}) d\lambda_2(\boldsymbol{A}) \cdots d\lambda_p(\boldsymbol{A}).$$

generalized Cauchy–Binet formula [22, Lemma 2], the integral in (47) can be evaluated, after some algebra, as

$$\mathcal{L}_0(\rho,\beta,N_{\rm c}) = \eta^{-\mathcal{T}_{\rm A}}\,\mathcal{T}_{\rm B}(\rho,N_{\rm c})\,\det\left(\begin{bmatrix}\mathsf{G}_{(n-m)}(\boldsymbol{\Phi}_2)\\ \boldsymbol{\Upsilon}(\rho,\beta)\end{bmatrix}\right), \quad N_{\rm c}\rho \neq 1,2,\ldots,m-1. \tag{48}$$

Substituting (48) into (17) gives the second case of (31). It should be noted that the formula in the second case of (31) has singular points at $N_{\rm c}\rho = 1,2,\ldots,m-1$ for each $\rho \in (0,1]$. These singularities stem from the quantity $\mathcal{T}_{\rm B}(\rho,N_{\rm c})$, which can be alleviated using the following analysis.

Suppose that $N_{\rm c}\rho$ is a positive integer. Then, using [23, Lemma 1], we have

$$\begin{aligned}\mathcal{L}_0(\rho,\beta,N_{\rm c}) &= \mathbb{E}_{\boldsymbol{\Theta}}\Big\{\mathbb{E}_{\boldsymbol{S}}\Big\{\mathrm{etr}\left(-\eta\boldsymbol{\Theta}\boldsymbol{S}\boldsymbol{S}^\dagger\right)\Big\}\Big\}\\ &= \mathbb{E}_{\boldsymbol{S}}\left\{\det\left(\boldsymbol{I}_{mn} + \eta\boldsymbol{S}\boldsymbol{S}^\dagger\boldsymbol{\Phi}_1\otimes\boldsymbol{\Phi}_2\right)^{-1}\right\}\\ &= \mathbb{E}_{\boldsymbol{S}}\left\{\det\left(\boldsymbol{I}_{mn} + \eta\boldsymbol{S}^\dagger\boldsymbol{\Phi}_1\boldsymbol{S}\otimes\boldsymbol{\Phi}_2\right)^{-1}\right\}\end{aligned} \tag{49}$$

where $\boldsymbol{S} \sim \tilde{\mathcal{N}}_{m,N_{\rm c}\rho}(\boldsymbol{0},\boldsymbol{I}_m,\boldsymbol{I}_{N_{\rm c}\rho})$ is a complex Gaussian matrix statistically independent of $\boldsymbol{\Theta}$, and the last equality follows from the fact that $\boldsymbol{S}\boldsymbol{S}^\dagger\boldsymbol{\Phi}_1$ and $\boldsymbol{S}^\dagger\boldsymbol{\Phi}_1\boldsymbol{S}$ have the same nonzero eigenvalues.

If $N_{\rm c}\rho \in \{1,2,\ldots,m\}$, then $\boldsymbol{Z} = \boldsymbol{S}^\dagger\boldsymbol{\Phi}_1\boldsymbol{S} \sim \tilde{\mathcal{Q}}_{N_{\rm c}\rho,m}(\boldsymbol{I}_m,\boldsymbol{I}_{N_{\rm c}\rho},\boldsymbol{\Phi}_1)$. Hence, using [23, Theorem 9], (49) for the case of $N_{\rm c}\rho \in \{1,2,\ldots,m\}$ can be written as

$$\begin{aligned}\mathcal{L}_0(\rho,\beta,N_{\rm c}) &= \mathbb{E}_{\boldsymbol{\lambda}(\boldsymbol{Z})}\left\{\prod_{k=1}^{N_{\rm c}\rho}\det\Big\{\boldsymbol{I}_n + \eta\lambda_k(\boldsymbol{Z})\boldsymbol{\Phi}_2\Big\}^{-1}\right\}\\ &= K_{\rm cor}^{-1}\int_{\boldsymbol{\lambda}(\boldsymbol{Z})}\prod_{k=1}^{N_{\rm c}\rho}\det\Big\{\boldsymbol{I}_n + \eta\lambda_k(\boldsymbol{Z})\boldsymbol{\Phi}_2\Big\}^{-1}\\ &\qquad\times\det\left(\begin{bmatrix}\mathsf{G}_{(m-N_{\rm c}\rho)}(\boldsymbol{\Phi}_1)\\ \grave{\boldsymbol{\Xi}}\end{bmatrix}\right)\det_{1\le i,j\le N_{\rm c}\rho}\left(\lambda_j^{i-1}(\boldsymbol{Z})\right)d\boldsymbol{\lambda}(\boldsymbol{Z})\end{aligned} \tag{50}$$

where $\grave{\boldsymbol{\Xi}} = \begin{bmatrix}\grave{\boldsymbol{\Xi}}_1 & \grave{\boldsymbol{\Xi}}_2 & \cdots & \grave{\boldsymbol{\Xi}}_{\varrho(\boldsymbol{Z})}\end{bmatrix}$ and the (i,j)th entry $\grave{\Xi}_{k,ij}$ of $\grave{\boldsymbol{\Xi}}_k \in \mathbb{R}^{N_{\rm c}\rho\times\chi_k(\boldsymbol{\Phi}_1)}$, $k=1,2,\ldots,\varrho(\boldsymbol{\Phi}_1)$, is given by

$$\grave{\Xi}_{k,ij} = \lambda_i^{j-1}(\boldsymbol{Z})e^{-\lambda_i(\boldsymbol{Z})/\lambda_{\langle k\rangle}(\boldsymbol{\Phi}_1)}. \tag{51}$$

Now, applying [22, Lemma 2] to (50) yields

$$\mathcal{L}_0(\rho,\beta,N_{\rm c}) = K_{\rm cor}^{-1}\,\det\left(\begin{bmatrix}\mathsf{G}_{(m-N_{\rm c}\rho)}(\boldsymbol{\Phi}_1)\\ \boldsymbol{\Xi}(\rho,\beta)\end{bmatrix}\right) \tag{52}$$

where the (i,j)th entry $\Xi_{k,ij}(\rho,\beta)$ of the kth constituent matrix $\boldsymbol{\Xi}_k(\rho,\beta)$ is given by

$$\Xi_{k,ij}(\rho,\beta)=\int_0^\infty \det\left(\boldsymbol{I}_n+\eta z\boldsymbol{\Phi}_2\right)^{-1} z^{i+j-2}e^{-z/\lambda_{\langle k\rangle}(\boldsymbol{\Phi}_1)}\,dz. \tag{53}$$

Using the *characteristic coefficients* [23, Definition 6], (53) can be written as

$$\Xi_{k,ij}(\rho,\beta)=\sum_{p=1}^{\varrho(\boldsymbol{\Phi}_2)}\sum_{q=1}^{\chi_p(\boldsymbol{\Phi}_2)}\mathcal{X}_{p,q}(\boldsymbol{\Phi}_2)\int_0^\infty\left(1+\eta\lambda_{\langle p\rangle}(\boldsymbol{\Phi}_2)z\right)^{-q}z^{i+j-2}e^{-z/\lambda_{\langle k\rangle}(\boldsymbol{\Phi}_1)}\,dz \tag{54}$$

where $\mathcal{X}_{p,q}(\boldsymbol{\Phi}_2)$ is the (p,q)th characteristic coefficient of $\boldsymbol{\Phi}_2$. Finally, substituting (52) into (17) gives the first case of (31) and hence, we complete the proof of the first part.

B. Proofs of Theorem 1.2 and 1.3

The second and third parts can be obtained by differentiating $\tilde{E}_0(\rho,\beta,N_\text{c})$ in Theorem 1.1 with respect to β and ρ, respectively, with the help of the logarithmic derivative of a determinant [28, Theorem 9.4] (or more generally [22, Lemma 1]).

V. Numerical Results and Discussion

In this section, we provide some numerical results to illustrate the reliability–rate tradeoff in block-fading MIMO channels. For spatial fading correlation, we consider an exponential correlation model with $\boldsymbol{\Phi}_\text{T}=\left(\zeta_\text{T}^{|i-j|}\right)$ and $\boldsymbol{\Phi}_\text{R}=\left(\zeta_\text{R}^{|i-j|}\right)$, $\zeta_\text{T},\,\zeta_\text{R}\in[0,1)$, in all examples.

To ascertain the effect of the channel coherence on the error exponent, Figs. 2 and 3, respectively, show the random coding exponent $E_\text{r}(R,N_\text{c})$ as a function of a rate R for i.i.d. ($\zeta_\text{T}=0$, $\zeta_\text{R}=0$) and exponentially correlated ($\zeta_\text{T}=0.5$, $\zeta_\text{R}=0.7$) MIMO channels at $\gamma=15$ dB, where $n_\text{T}=n_\text{R}=3$ and N_c ranges from 1 to 10. We can see from the figures that the exponent at a rate R below the ergodic capacity decreases with N_c, while the ergodic capacity remains constant for all N_c (i.e., 8.48 and 7.19 nats/symbol for Figs. 2 and 3, respectively). For example, the error exponents at rates $R\le R_\text{cr}$ for $N_\text{c}=10$ are reduced by roughly 3.46 and 2.86 for i.i.d. and exponentially correlated cases, respectively, compared with those for $N_\text{c}=1$. This reduction in the exponent, consequently, requires using a longer code to achieve the same error probability. Hence, we see that unlike the capacity (with perfect receive CSI), the channel coherence time plays a fundamental role in the error exponent or the reliability of communications.

Fig. 4 demonstrates the effect of spatial fading correlation on the random coding exponent, where $\zeta_\text{T}=\zeta_\text{R}=\zeta$, $\gamma=15$ dB, $n_\text{T}=n_\text{R}=3$, $N_\text{c}=5$, and ζ ranges from 0 (i.i.d.) to 0.9. As

seen from the figure, there exists a remarkable reduction in the exponent at the same rate due to correlation, especially for $\zeta \geq 0.5$. The amount of reduction in the exponent at rates $R \leq R_{\mathrm{cr}}$, relative to the i.i.d. MIMO exponent, ranges from 0.07 for $\zeta = 0.2$ to 2.17 for $\zeta = 0.9$, indicating that a longer code is required to achieve the same level of reliable communications. Equivalently, a decrease in the information rate is required for more correlated channels to achieve the same value of the exponent. For example, the exponent at a rate 3 nats/symbol are 1.94 and 1.53 for the i.i.d. and correlated ($\zeta_{\mathrm{T}} = \zeta_{\mathrm{R}} = 0.5$) channels, respectively. This implies that 27% increase in the codeword length, due to spatial fading correlation, is required to achieve a rate 3 nats/symbol with the same communication reliability.

To get more insight into the influences of the number of antennas, channel coherence time, and fading correlation on a coding requirement for MIMO channels, the codeword length required to achieve $P_{\mathrm{e}} \leq 10^{-6}$ at a rate 8.0 bits/symbol (5.55 nats/symbol) are investigated in Tables II–IV. The codeword lengths in the tables are calculated in such a manner as described in Section III-C. Table II serves to demonstrate the effect of increasing the number of antennas on the coding requirement, in which the required codeword length L is shown for i.i.d. MIMO channels with $N_{\mathrm{c}} = 5$. Note that it is impossible to reliably communicate at a rate 8.0 bits/symbol below the SNR γ of 14.55 dB, 9.68 dB, and 6.79 dB for $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$, 3, and 4, respectively, since these SNR's are required to attain the ergodic capacity $\langle C \rangle$ of 8.0 bits/symbol in each of the cases. As seen from the table, with increasing the number of antennas at both transmit and receive sides, the required codeword lengths are remarkably reduced. This is due to the advantages of the use of multiple antennas, e.g., spatial multiplexing and diversity gains [21]. For example, at $\gamma = 16$ dB, increasing the number of antennas at both sides from 2 to 3 and 4 reduces the corresponding codeword length to almost 2.8% and 0.9% of the amount required for two transmit and receive antennas, respectively, which is a tremendous reduction in the codeword length.

Table III shows the required codeword length L for i.i.d. and exponentially correlated ($\zeta_{\mathrm{T}} = 0.5$, $\zeta_{\mathrm{R}} = 0.7$) MIMO channels with $n_{\mathrm{T}} = n_{\mathrm{R}} = 3$ at $\gamma = 15$ dB when N_{c} varies from 1 to 10. It is clear from Table III that for each value of N_{c}, the codeword lengths for correlated channels are much longer than those for i.i.d. channels. For example, the increase in the required codeword length, due to exponential correlation ($\zeta_{\mathrm{T}} = 0.5$, $\zeta_{\mathrm{R}} = 0.7$), ranges from 194% for $N_{\mathrm{c}} = 1$ to 138% for $N_{\mathrm{c}} = 10$, which is a significant increase in required codeword length. Also, when going N_{c} from 1 to 10, there is a considerable increase in the required codeword length, relative

to that for the single-symbol coherence time, which ranges from 33% to 344% for the i.i.d. case and from 28% to 258% for the correlated case, respectively.

Table IV demonstrates the effect of correlation on the required code length L, where $n_{\rm T} = n_{\rm R} = 3$, $\zeta_{\rm T} = \zeta_{\rm R} = \zeta$, $N_{\rm c} = 5$, and $\gamma = 15$ dB. The table contains the corresponding codeword lengths for ζ from 0 to 0.9. As seen from the table, the required codeword length for the case of exponential correlation $\zeta = 0.7$ is equal to 4.5 times as long as for the i.i.d. channel ($\zeta = 0$). Particularly, when $\zeta \geq 0.5$, there exists a large amount of increase in required codeword length due to a stronger correlation. Also, since the ergodic capacity is 7.36 bits/symbol for $\zeta_{\rm T} = \zeta_{\rm R} = 0.9$ at $\gamma = 15$ dB, it is impossible to achieve reliable communications at a rate 8.0 bits/symbol (regardless of the codeword length), when $\zeta_{\rm T} = \zeta_{\rm R} = 0.9$.

Finally, Fig. 5 shows the cutoff rate R_0 in nats/symbol as a function of a correlation coefficient ζ for exponentially correlated MIMO channels with $\zeta_{\rm T} = \zeta_{\rm R} = \zeta$ at $\gamma = 15$ dB, where $n_{\rm R} = n_{\rm R} = 3$ and $N_{\rm c}$ varies from 1 to 10. We see that the cutoff rate R_0 decreases with $N_{\rm c}$ for all $\zeta \in [0, 1)$. While $\langle C \rangle$ remains constant, R_0 monotonically decreases with $N_{\rm c}$, going to 0 as $N_{\rm c} \to \infty$ (see (29) and (30) with $\rho = 1$ and $\beta = n_{\rm T}$). Hence, these two measures diverge as $N_{\rm c}$ increases and eventually $\lim_{N_{\rm c} \to \infty} \frac{\langle C \rangle}{R_0} = \infty$, which coincides with the divergent behavior of the capacity and cutoff rate of a channel with block memory [9]. This observation reveals that R_0 is more pertinent than $\langle C \rangle$ as a figure of merit that reflects the quality of block-fading channels.

VI. Conclusions

In this paper, we derived Gallager's random coding error exponent to investigate a fundamental tradeoff between the communication reliability and information rate in spatially correlated MIMO channels. We considered a block-fading channel with perfect receive CSI and Gaussian codebooks. The required codeword lengths for a prescribed error probability were calculated from the random coding bound to aid in the assessment of the coding requirement on such MIMO channels, taking into account the effects of the number of antennas, the channel coherence time, and the amount of spatial fading correlation. In addition, we obtained the general formulae for the ergodic capacity and cutoff rate, which encompass all the previous capacity results as special cases of our expressions. In parallel to the capacity–cutoff rate divergence in a block-memory channel, we observed the channel-incurable effect: the monotonically decreasing property of the MIMO exponent (i.e., communication reliability) with the channel coherence time.

APPENDIX

A. Proof of Proposition 1

Lemma 1: Let $\boldsymbol{S} \sim \tilde{\mathcal{N}}_{m,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$ and $\boldsymbol{A} \in \mathbb{C}^{m\times m} > 0$ be Hermitian. Then, we have

$$\mathbb{E}\left\{\operatorname{etr}\left(-\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\dagger\right)\right\} = \det\left(\boldsymbol{I}_m + \boldsymbol{\Sigma}\boldsymbol{A}\right)^{-n} \operatorname{etr}\left\{-\left(\boldsymbol{A}^{-1} + \boldsymbol{\Sigma}\right)^{-1} \boldsymbol{M}\boldsymbol{M}^\dagger\right\}. \tag{55}$$

Proof: Note that

$$\mathbb{E}\left\{\operatorname{etr}\left(-\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\dagger\right)\right\} = \frac{\det\left(\boldsymbol{\Sigma}\right)^{-n}}{\pi^{mn}} \int_{\boldsymbol{S}} \operatorname{etr}\left\{-\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\dagger - \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{S}-\boldsymbol{M}\right)\left(\boldsymbol{S}-\boldsymbol{M}\right)^\dagger\right\} d\boldsymbol{S}. \tag{56}$$

By writing the trace of the quadratic form in the exponent of (56) as

$$\begin{aligned}
&\operatorname{tr}\left\{\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\dagger + \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{S}-\boldsymbol{M}\right)\left(\boldsymbol{S}-\boldsymbol{M}\right)^\dagger\right\} \\
&\quad = \operatorname{tr}\left\{\left(\boldsymbol{A}+\boldsymbol{\Sigma}^{-1}\right)\left[\boldsymbol{S} - \left(\boldsymbol{I}_m + \boldsymbol{\Sigma}\boldsymbol{A}\right)^{-1}\boldsymbol{M}\right]\left[\boldsymbol{S} - \left(\boldsymbol{I}_m + \boldsymbol{\Sigma}\boldsymbol{A}\right)^{-1}\boldsymbol{M}\right]^\dagger - \left(\boldsymbol{A}^{-1}+\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{M}\boldsymbol{M}^\dagger\right\},
\end{aligned} \tag{57}$$

we get

$$\begin{aligned}
\mathbb{E}\left\{\operatorname{etr}\left(-\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\dagger\right)\right\} &= \frac{\det\left(\boldsymbol{\Sigma}\right)^{-n}}{\pi^{mn}} \operatorname{etr}\left\{-\left(\boldsymbol{A}^{-1}+\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{M}\boldsymbol{M}^\dagger\right\} \\
&\quad \times \underbrace{\int_{\boldsymbol{S}} \operatorname{etr}\left\{-\left(\boldsymbol{A}+\boldsymbol{\Sigma}^{-1}\right)\left[\boldsymbol{S} - \left(\boldsymbol{I}_m + \boldsymbol{\Sigma}\boldsymbol{A}\right)^{-1}\boldsymbol{M}\right]\left[\boldsymbol{S} - \left(\boldsymbol{I}_m + \boldsymbol{\Sigma}\boldsymbol{A}\right)^{-1}\boldsymbol{M}\right]^\dagger\right\} d\boldsymbol{S}}_{=\pi^{mn}\det\left(\boldsymbol{A}+\boldsymbol{\Sigma}^{-1}\right)^{-n}}
\end{aligned} \tag{58}$$

from which (55) follows readily. □

Proof of Proposition 1: Using Lemma 1, we have

$$\begin{aligned}
&\int_{\boldsymbol{X}} p_{\boldsymbol{X}}\left(\boldsymbol{X}\right) e^{r\left[\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\dagger\right) - N_{\mathrm{c}}\mathcal{P}\right]} p\left(\boldsymbol{Y} \middle| \boldsymbol{X}, \boldsymbol{H}\right)^{1/(1+\rho)} d\boldsymbol{X} \\
&\quad = e^{-rN_{\mathrm{c}}\mathcal{P}} \left(\pi N_0\right)^{-n_{\mathrm{R}}N_{\mathrm{c}}/(1+\rho)} \det\left(\boldsymbol{I}_{n_{\mathrm{T}}} - r\boldsymbol{Q}\right)^{-N_{\mathrm{c}}} \det\left(\boldsymbol{I}_{n_{\mathrm{R}}} + \frac{\boldsymbol{H}\left(\boldsymbol{Q}^{-1} - r\boldsymbol{I}_{n_{\mathrm{T}}}\right)^{-1}\boldsymbol{H}^\dagger}{N_0\left(1+\rho\right)}\right)^{-N_{\mathrm{c}}} \\
&\qquad \times \operatorname{etr}\left\{-\frac{1}{N_0\left(1+\rho\right)}\left(\boldsymbol{I}_{n_{\mathrm{R}}} + \frac{\boldsymbol{H}\left(\boldsymbol{Q}^{-1} - r\boldsymbol{I}_{n_{\mathrm{T}}}\right)^{-1}\boldsymbol{H}^\dagger}{N_0\left(1+\rho\right)}\right)^{-1}\boldsymbol{Y}\boldsymbol{Y}^\dagger\right\}.
\end{aligned} \tag{59}$$

Substituting (59) into (13) and integrating over $\boldsymbol{Y}$, we have

$$\begin{aligned}
&\int_{\boldsymbol{Y}} \left\{\int_{\boldsymbol{X}} p_{\boldsymbol{X}}\left(\boldsymbol{X}\right) e^{r\left[\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\dagger\right) - N_{\mathrm{c}}\mathcal{P}\right]} p\left(\boldsymbol{Y} \middle| \boldsymbol{X}, \boldsymbol{H}\right)^{1/(1+\rho)} d\boldsymbol{X}\right\}^{1+\rho} d\boldsymbol{Y} \\
&\quad = e^{-rN_{\mathrm{c}}\mathcal{P}(1+\rho)} \det\left(\boldsymbol{I}_{n_{\mathrm{T}}} - r\boldsymbol{Q}\right)^{-N_{\mathrm{c}}(1+\rho)} \det\left(\boldsymbol{I}_{n_{\mathrm{R}}} + \frac{\boldsymbol{H}\left(\boldsymbol{Q}^{-1} - r\boldsymbol{I}_{n_{\mathrm{T}}}\right)^{-1}\boldsymbol{H}^\dagger}{N_0\left(1+\rho\right)}\right)^{-N_{\mathrm{c}}\rho}.
\end{aligned} \tag{60}$$

Finally, substituting (60) into (13) yields the result (15).

B. Proof of Proposition 2

We provide a sketch of the proof of Proposition 2 using a similar approach in [4] and [6]. For notational simplicity, let us denote $\mathbf{\Omega}_{\rho,\beta} = \beta \boldsymbol{I}_m + \gamma \mathbf{\Theta} / (1+\rho)$. Then, $\tilde{E}_0(\rho, \beta, N_\text{c})$ in (17) can be rewritten as

$$\tilde{E}_0(\rho, \beta, N_\text{c}) = \mathcal{K}(\rho, \beta) - m\rho \ln(\beta) - \frac{1}{N_\text{c}} \ln \mathcal{L}_1(\rho, \beta, N_\text{c}) \tag{61}$$

where $\mathcal{L}_1(\rho, \beta, N_\text{c}) = \mathbb{E}\big\{\det(\mathbf{\Omega}_{\rho,\beta})^{-N_\text{c}\rho}\big\}$. Since $\mathcal{K}(\rho, \beta) - m\rho\ln(\beta)$ is concave in β, $\tilde{E}_0(\rho, \beta, N_\text{c})$ is a concave function of β if $\ln \mathcal{L}_1^{-1}(\rho, \beta, N_\text{c})$ is concave in β for all $\rho \in [0,1]$. Noting that

$$\frac{\partial^2 \ln \mathcal{L}_1^{-1}(\rho, \beta, N_\text{c})}{\partial \beta^2} = \mathcal{L}_1^{-2}(\rho, \beta, N_\text{c}) \left\{ \left(\frac{\partial \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta} \right)^2 - \mathcal{L}_1(\rho, \beta, N_\text{c}) \frac{\partial^2 \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta^2} \right\} \tag{62}$$

and $\mathcal{L}_1(\rho, \beta, N_\text{c}) \geq 0$, it is sufficient to show that

$$\left(\frac{\partial \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta} \right)^2 \leq \mathcal{L}_1(\rho, \beta, N_\text{c}) \frac{\partial^2 \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta^2}. \tag{63}$$

It is easy to show that

$$\frac{\partial \det(\mathbf{\Omega}_{\rho,\beta})}{\partial \beta} = \det(\mathbf{\Omega}_{\rho,\beta})\, \operatorname{tr}\left(\mathbf{\Omega}_{\rho,\beta}^{-1}\right) \quad \text{and} \quad \frac{\partial \operatorname{tr}\left(\mathbf{\Omega}_{\rho,\beta}^{-1}\right)}{\partial \beta} = -\operatorname{tr}\left(\mathbf{\Omega}_{\rho,\beta}^{-2}\right)$$

and hence,

$$\frac{\partial \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta} = \mathbb{E}\left\{ -N_\text{c}\rho \det(\mathbf{\Omega}_{\rho,\beta})^{-N_\text{c}\rho} \operatorname{tr}\left(\mathbf{\Omega}_{\rho,\beta}^{-1}\right) \right\} \tag{64}$$

$$\frac{\partial^2 \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta^2} = \mathbb{E}\left\{ N_\text{c}\rho \det(\mathbf{\Omega}_{\rho,\beta})^{-N_\text{c}\rho} \left[N_\text{c}\rho \operatorname{tr}^2\left(\mathbf{\Omega}_{\rho,\beta}^{-1}\right) + \operatorname{tr}\left(\mathbf{\Omega}_{\rho,\beta}^{-2}\right) \right] \right\}. \tag{65}$$

Let us now define the random variables

$$\mathsf{X}^2 = \det(\mathbf{\Omega}_{\rho,\beta})^{-N_\text{c}\rho} \quad \text{and} \quad \mathsf{Y}^2 = (N_\text{c}\rho)^2 \det(\mathbf{\Omega}_{\rho,\beta})^{-N_\text{c}\rho} \operatorname{tr}^2\left(\mathbf{\Omega}_{\rho,\beta}^{-1}\right).$$

From Schwartz's inequality, we have

$$\begin{aligned} \left(\frac{\partial \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta} \right)^2 &= \mathbb{E}^2\{\mathsf{XY}\} \leq \mathbb{E}\left\{\mathsf{X}^2\right\} \cdot \mathbb{E}\left\{\mathsf{Y}^2\right\} \\ &\leq \mathbb{E}\left\{\mathsf{X}^2\right\} \cdot \mathbb{E}\left\{\mathsf{Y}^2 + N_\text{c}\rho \mathsf{X}^2 \operatorname{tr}\left(\mathbf{\Omega}_{\rho,\beta}^{-2}\right)\right\} \\ &= \mathcal{L}_1(\rho, \beta, N_\text{c}) \frac{\partial^2 \mathcal{L}_1(\rho, \beta, N_\text{c})}{\partial \beta^2}. \end{aligned} \tag{66}$$

From (66), we see that $\tilde{E}_0(\rho, \beta, N_\text{c})$ is a concave function of β for all $\rho \in [0,1]$. Hence, the maximum over β occurs at $\beta^*(\rho)$ for which $\left[\partial \tilde{E}_0(\rho, \beta, N_\text{c}) / \partial \beta\right]\big|_{\beta=\beta^*(\rho)} = 0$ and it is sufficient

to show that $\left[\partial \tilde{E}_0(\rho,\beta,N_{\mathrm{c}})/\partial\beta\right]\big|_{\beta=0} \geq 0$ and $\left[\partial \tilde{E}_0(\rho,\beta,N_{\mathrm{c}})/\partial\beta\right]\big|_{\beta=n_{\mathrm{T}}} \leq 0$ for all $\rho \in [0,1]$ in order to prove $0 < \beta^*(\rho) \leq n_{\mathrm{T}}$. Since

$$\frac{\partial \tilde{E}_0(\rho,\beta,N_{\mathrm{c}})}{\partial\beta} = \underbrace{\frac{(1+\rho)(n_{\mathrm{T}}-\beta)}{\beta}}_{\triangleq \mathcal{K}^{(\beta)}(\rho,\beta)=\frac{\partial\mathcal{K}(\rho,\beta)}{\partial\beta}} - \frac{m\rho}{\beta} - \frac{1}{N_{\mathrm{c}}}\mathcal{L}_1^{-1}(\rho,\beta,N_{\mathrm{c}})\frac{\partial\mathcal{L}_1(\rho,\beta,N_{\mathrm{c}})}{\partial\beta}, \tag{67}$$

it is clear that $\lim_{\beta\to 0}\partial\tilde{E}_0(\rho,\beta,N_{\mathrm{c}})/\partial\beta \to \infty$. Also,

$$\left[\frac{\partial \tilde{E}_0(\rho,\beta,N_{\mathrm{c}})}{\partial\beta}\right]\Bigg|_{\beta=n_{\mathrm{T}}} = -\frac{m\rho}{n_{\mathrm{T}}} - \frac{1}{N_{\mathrm{c}}}\mathcal{L}_1^{-1}(\rho,n_{\mathrm{T}},N_{\mathrm{c}})\left[\frac{\partial\mathcal{L}_1(\rho,\beta,N_{\mathrm{c}})}{\partial\beta}\right]\Bigg|_{\beta=n_{\mathrm{T}}}. \tag{68}$$

Since $\mathcal{L}_1(\rho,\beta,N_{\mathrm{c}}) \geq 0$ and

$$\begin{aligned}\mathcal{L}_1(\rho,n_{\mathrm{T}},N_{\mathrm{c}}) &= \mathbb{E}\left\{\det\left(\mathbf{\Omega}_{\rho,n_{\mathrm{T}}}\right)^{-N_{\mathrm{c}}\rho}\right\} = \mathbb{E}\left\{\det\left(\mathbf{\Omega}_{\rho,n_{\mathrm{T}}}\right)^{-N_{\mathrm{c}}\rho}\ \frac{\operatorname{tr}\left(\mathbf{\Omega}_{\rho,n_{\mathrm{T}}}^{-1}\right)}{\operatorname{tr}\left(\mathbf{\Omega}_{\rho,n_{\mathrm{T}}}^{-1}\right)}\right\} \\ &\geq \frac{n_{\mathrm{T}}}{m}\,\mathbb{E}\left\{\det\left(\mathbf{\Omega}_{\rho,n_{\mathrm{T}}}\right)^{-N_{\mathrm{c}}\rho}\ \operatorname{tr}\left(\mathbf{\Omega}_{\rho,n_{\mathrm{T}}}^{-1}\right)\right\} \\ &= -\frac{n_{\mathrm{T}}}{m\rho}\cdot\frac{1}{N_{\mathrm{c}}}\left[\frac{\partial\mathcal{L}_1(\rho,\beta,N_{\mathrm{c}})}{\partial\beta}\right]\Bigg|_{\beta=n_{\mathrm{T}}},\end{aligned} \tag{69}$$

it follows that

$$-\frac{m\rho}{n_{\mathrm{T}}} - \frac{1}{N_{\mathrm{c}}}\mathcal{L}_1^{-1}(\rho,n_{\mathrm{T}},N_{\mathrm{c}})\left[\frac{\partial\mathcal{L}_1(\rho,\beta,N_{\mathrm{c}})}{\partial\beta}\right]\Bigg|_{\beta=n_{\mathrm{T}}} \leq 0. \tag{70}$$

Thus, $\left[\partial \tilde{E}_0(\rho,\beta,N_{\mathrm{c}})/\partial\beta\right]\big|_{\beta=n_{\mathrm{T}}} \leq 0$ and we complete the proof of the proposition.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July–Oct. 1948.

[2] ——, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, May 1959.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[4] T. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inform. Theory*, vol. 16, no. 3, pp. 353–355, May 1970.

[5] W. K. M. Ahmed, "Information theoretic reliability function for flat fading channels," Ph.D. dissertation, Queen's University, Kingston, ON, Canada, Sept. 1997.

[6] W. K. M. Ahmed and P. J. McLane, "Random coding error exponents for two-dimensional flat fading channels with complete channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 4, pp. 1338–1346, May 1999.

[7] I. Abou-Faycal and B. M. Hochwald, "Coding requirements for multiple-antenna channels with unknown Rayleigh fading," *Tech. Memo.*, Bell Laboratories, Lucent Technologies, 1999.

[8] G. Kaplan and S. Shamai (Shitz), "Error probabilities for the block-fading Gaussian channel," *Arch. Electronik Übertragungstech.*, vol. 49, no. 4, pp. 192–205, 1995.

[9] R. J. McEliece and W. E. Stark, "Channel with block interference," *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 44–53, Jan. 1984.

[10] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[11] B. M. Hochwald, T. L. Marzetta, and B. Hassibi, "Space–time autocoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2761–2781, Nov. 2001.

[12] J. H. Winters, "On the capacity of radio communication systems with diversity in Rayleigh fading environment," *IEEE J. Select. Areas Commun.*, vol. 5, no. 5, pp. 871–878, June 1987.

[13] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication system," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1740–1751, Feb./Mar./Apr. 1994.

[14] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[15] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[16] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[17] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[18] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[19] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[20] P. J. Smith, S. Roy, and M. Shafi, "Capacity of MIMO systems with semicorrelated flat fading," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2781–2788, Oct. 2003.

[21] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.

[22] H. Shin, M. Z. Win, J. H. Lee, and M. Chiani, "On the capacity of doubly correlated MIMO channels," *IEEE Trans. Wireless Commun.*, to be published. [Online]. Available: http://arxiv.org/abs/cs.IT/0509075

[23] H. Shin and M. Z. Win, "MIMO diversity in the presence of double scattering," *IEEE Trans. Inform. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/cs.IT/0511028

[24] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statistics*, vol. 35, no. 2, pp. 475–501, June 1964.

[25] C. G. Khatri, "On certain distribution problems based on positive definite quadratic functions in normal vectors," *Ann. Math. Statistics*, vol. 37, no. 2, pp. 468–479, Apr. 1966.

[26] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[27] H. Shin, "Capacity and error exponents for multiple-input multiple-output wireless channels," Ph.D. dissertation, Seoul National University, Seoul, Korea, Aug. 2004.

[28] P. D. Lax, *Linear Algebra*. New York: Wiley, 1997.

[29] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic, 2000.

TABLE I

SOME QUANTITIES AND MATRICES INVOLVED IN THEOREM 1

In Theorem 1

$$\mathsf{G}_{(\nu)}(\boldsymbol{\Psi}) = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1 & \boldsymbol{\mathcal{A}}_2 & \cdots & \boldsymbol{\mathcal{A}}_{\varrho(\boldsymbol{\Psi})} \end{bmatrix} \in \mathbb{R}^{\nu\times p}, \qquad \text{for } \boldsymbol{\Psi}\ p\times p \text{ Hermitian}, \ \nu \le p$$

$$\overline{\mathsf{G}}_{(\nu)}(\boldsymbol{\Psi}) = \begin{bmatrix} \overline{\boldsymbol{\mathcal{A}}}_1 & \overline{\boldsymbol{\mathcal{A}}}_2 & \cdots & \overline{\boldsymbol{\mathcal{A}}}_{\varrho(\boldsymbol{\Psi})} \end{bmatrix} \in \mathbb{R}^{\nu\times p}$$

where[1)]

$$\boldsymbol{\mathcal{A}}_k = (\mathcal{A}_{k,ij}) \in \mathbb{R}^{\nu\times\chi_k(\boldsymbol{\Psi})},\ k=1,2,\ldots,\varrho(\boldsymbol{\Psi})$$

$$(i,j)^{\underline{\text{th}}} \text{ element:} \quad \mathcal{A}_{k,ij} = (-1)^{i-j}(i-j+1)_{j-1}\,\lambda_{\langle k\rangle}^{-i+j}(\boldsymbol{\Psi})$$

$$\overline{\boldsymbol{\mathcal{A}}}_k = (\overline{\mathcal{A}}_{k,ij}) \in \mathbb{R}^{\nu\times\chi_k(\boldsymbol{\Psi})},\ k=1,2,\ldots,\varrho(\boldsymbol{\Psi})$$

$$(i,j)^{\underline{\text{th}}} \text{ element:} \quad \overline{\mathcal{A}}_{k,ij} = (-1)^{i-j}(i-j+1)_{j-1}\,\lambda_{\langle k\rangle}^{i-j}(\boldsymbol{\Psi}).$$

[1)] $(a)_n = a(a+1)\cdots(a+n-1)$, $(a)_0 = 1$ is the Pochhammer symbol.

$$K_{\text{cor}} = \det(\boldsymbol{\Phi}_1)^{N_c\rho}\det\left\{\mathsf{G}_{(m)}(\boldsymbol{\Phi}_1)\right\}\prod_{k=1}^{N_c\rho}(k-1)!\,, \qquad N_c\rho\in\{1,2,\ldots,m\}$$

$$K_{\text{iid}} = \prod_{k=1}^{m}(n-k)!\,(k-1)!$$

$$\mathcal{T}_{\text{A}} = \tfrac{1}{2}m(m+1) - \tfrac{1}{2}\sum_{i=1}^{\varrho(\boldsymbol{\Phi}_1)}\chi_i(\boldsymbol{\Phi}_1)\Big[\chi_i(\boldsymbol{\Phi}_1)+1\Big]$$

$$\mathcal{T}_{\text{B}}(\rho,N_c) = \det(\boldsymbol{\Phi}_2)^{-m}\det\left\{\overline{\mathsf{G}}_{(m)}(\boldsymbol{\Phi}_1)\right\}^{-1}\det\left\{\mathsf{G}_{(n)}(\boldsymbol{\Phi}_2)\right\}^{-1}\frac{\prod_{i=1}^{\varrho(\boldsymbol{\Phi}_1)}\prod_{j=1}^{\chi_i(\boldsymbol{\Phi}_1)}(N_c\rho-m+1)_{j-1}}{\prod_{k=1}^{m}(N_c\rho-m+1)_{k-1}}$$

$$\boldsymbol{\Xi}(\rho,\beta) = \Big[\boldsymbol{\Xi}_1(\rho,\beta)\quad \boldsymbol{\Xi}_2(\rho,\beta)\quad\cdots\quad \boldsymbol{\Xi}_{\varrho(\boldsymbol{\Phi}_1)}(\rho,\beta)\Big] \in \mathbb{R}^{N_c\rho\times m}$$

where[2)]

$$\boldsymbol{\Xi}_k(\rho,\beta) = (\Xi_{k,ij}(\rho,\beta)) \in \mathbb{R}^{N_c\rho\times\chi_k(\boldsymbol{\Phi}_1)},\ k=1,2,\ldots,\varrho(\boldsymbol{\Phi}_1),\ N_c\rho\in\{1,2,\ldots,m\}$$

$(i,j)^{\underline{\text{th}}}$ element:

$$\Xi_{k,ij}(\rho,\beta) = \sum_{p=1}^{\varrho(\boldsymbol{\Phi}_2)}\sum_{q=1}^{\chi_p(\boldsymbol{\Phi}_2)}\mathcal{X}_{p,q}(\boldsymbol{\Phi}_2)\,\mathcal{G}_{i+j-1,1}\left(\tfrac{\gamma}{\beta(1+\rho)}\lambda_{\langle p\rangle}(\boldsymbol{\Phi}_2),\lambda_{\langle k\rangle}(\boldsymbol{\Phi}_1),-q+1\right).$$

[2)] $\mathcal{X}_{p,q}(\boldsymbol{\Phi}_2)$ is the (p,q)th *characteristic coefficient* of $\boldsymbol{\Phi}_2$ (see for details [23, Definition 6]).

The $\mathcal{G}_{\kappa,\nu}(a,b,\mu)$ is defined as the integral

$$\begin{aligned}\mathcal{G}_{\kappa,\nu}(a,b,\mu) &= \int_0^\infty (1+ax)^{\mu-1}\ln^{\nu-1}(1+ax)\,x^{\kappa-1}e^{-x/b}dx, \qquad a,b>0,\ \kappa,\nu\in\mathbb{N}_b,\ \mu\in\mathbb{C}\\ &= \begin{cases} b^{\kappa}(\kappa-1)!\,{}_2F_0(\kappa,-\mu+1;-ab)\,, & \text{if } \nu=1\\ a^{-\kappa}(\nu-1)!\,e^{1/(ab)}\sum_{k=0}^{\kappa-1}\Big[(-1)^{\kappa-k-1}\binom{\kappa-1}{k}(ab)^{\mu+k}G_{\nu,\nu+1}^{\nu+1,0}\Big(\frac{1}{ab}\Big|\begin{smallmatrix}1,1,\ldots,1\\0,0,\ldots,0,\mu+k\end{smallmatrix}\Big)\Big], & \text{otherwise}\end{cases}\end{aligned}$$

where ${}_pF_q(a_1,a_2,\ldots,a_p;b_1,b_2,\ldots,b_q;z)$ is the generalized hypergeometric function of scalar argument [29, eq. (9.14.1)] and $G_{p,q}^{m,n}(\cdot)$ is the Meijer G-function [29, eq. (9.301)]. The detailed derivation of this integral identity can be found in [22, Appendix A].

TABLE I

(Continued.) SOME QUANTITIES AND MATRICES INVOLVED IN THEOREM 1

In Theorem 1

$$\mathbf{\Upsilon}(\rho,\beta) = \begin{bmatrix} \mathbf{\Upsilon}_{1,1}(\rho,\beta) & \cdots & \mathbf{\Upsilon}_{1,\varrho(\mathbf{\Phi}_2)}(\rho,\beta) \\ \vdots & \ddots & \vdots \\ \mathbf{\Upsilon}_{\varrho(\mathbf{\Phi}_1),1}(\rho,\beta) & \cdots & \mathbf{\Upsilon}_{\varrho(\mathbf{\Phi}_1),\varrho(\mathbf{\Phi}_2)}(\rho,\beta) \end{bmatrix} \in \mathbb{R}^{m\times n}$$

where

$\mathbf{\Upsilon}_{p,q}(\rho,\beta) = \left(\Upsilon_{p,q,ij}(\rho,\beta)\right) \in \mathbb{R}^{\chi_p(\mathbf{\Phi}_1)\times\chi_q(\mathbf{\Phi}_2)}$, $p = 1,2,\ldots,\varrho(\mathbf{\Phi}_1)$, $q = 1,2,\ldots,\varrho(\mathbf{\Phi}_2)$

$(i,j)^{\underline{\text{th}}}$ element: $\Upsilon_{p,q,ij}(\rho,\beta) = \mathcal{G}_{i+j-1,1}\left(\frac{\gamma}{\beta(1+\rho)}\lambda_{\langle p\rangle}(\mathbf{\Phi}_1), \lambda_{\langle q\rangle}(\mathbf{\Phi}_2), -N_{\text{c}}\rho + m - i + 1\right)$.

$$\mathbf{\Upsilon}^{(\beta)}(\rho,\beta) \triangleq \frac{\partial}{\partial\beta}\mathbf{\Upsilon}(\rho,\beta) = \begin{bmatrix} \mathbf{\Upsilon}^{(\beta)}_{1,1}(\rho,\beta) & \cdots & \mathbf{\Upsilon}^{(\beta)}_{1,\varrho(\mathbf{\Phi}_2)}(\rho,\beta) \\ \vdots & \ddots & \vdots \\ \mathbf{\Upsilon}^{(\beta)}_{\varrho(\mathbf{\Phi}_1),1}(\rho,\beta) & \cdots & \mathbf{\Upsilon}^{(\beta)}_{\varrho(\mathbf{\Phi}_1),\varrho(\mathbf{\Phi}_2)}(\rho,\beta) \end{bmatrix}$$

where

$\mathbf{\Upsilon}^{(\beta)}_{p,q}(\rho,\beta) = \left(\Upsilon^{(\beta)}_{p,q,ij}(\rho,\beta)\right) \in \mathbb{R}^{\chi_p(\mathbf{\Phi}_1)\times\chi_q(\mathbf{\Phi}_2)}$

$(i,j)^{\underline{\text{th}}}$ element:

$\Upsilon^{(\beta)}_{p,q,ij}(\rho,\beta) = \frac{\gamma}{\beta^2(1+\rho)}\left(N_{\text{c}}\rho - m + i\right)\lambda_{\langle p\rangle}(\mathbf{\Phi}_1)\,\mathcal{G}_{i+j,1}\left(\frac{\gamma}{\beta(1+\rho)}\lambda_{\langle p\rangle}(\mathbf{\Phi}_1), \lambda_{\langle q\rangle}(\mathbf{\Phi}_2), -N_{\text{c}}\rho + m - i\right)$.

$$\mathbf{\Upsilon}^{(\rho)}(\rho,\beta) \triangleq \frac{\partial}{\partial\rho}\mathbf{\Upsilon}(\rho,\beta) = \begin{bmatrix} \mathbf{\Upsilon}^{(\rho)}_{1,1}(\rho,\beta) & \cdots & \mathbf{\Upsilon}^{(\rho)}_{1,\varrho(\mathbf{\Phi}_2)}(\rho,\beta) \\ \vdots & \ddots & \vdots \\ \mathbf{\Upsilon}^{(\rho)}_{\varrho(\mathbf{\Phi}_1),1}(\rho,\beta) & \cdots & \mathbf{\Upsilon}^{(\rho)}_{\varrho(\mathbf{\Phi}_1),\varrho(\mathbf{\Phi}_2)}(\rho,\beta) \end{bmatrix}$$

where

$\mathbf{\Upsilon}^{(\rho)}_{p,q}(\rho,\beta) = \left(\Upsilon^{(\rho)}_{p,q,ij}(\rho,\beta)\right) \in \mathbb{R}^{\chi_p(\mathbf{\Phi}_1)\times\chi_q(\mathbf{\Phi}_2)}$

$(i,j)^{\underline{\text{th}}}$ element:

$\Upsilon^{(\rho)}_{p,q,ij}(\rho,\beta) = \frac{\beta}{1+\rho}\Upsilon^{(\beta)}_{p,q,ij}(\rho,\beta) - N_{\text{c}}\,\mathcal{G}_{i+j-1,2}\left(\frac{\gamma}{\beta(1+\rho)}\lambda_{\langle p\rangle}(\mathbf{\Phi}_1), \lambda_{\langle q\rangle}(\mathbf{\Phi}_2), -N_{\text{c}}\rho + m - i + 1\right)$.

$\mathbf{\Upsilon}_{\text{iid}}(\rho,\beta) = \left(\Upsilon_{\text{iid},ij}(\rho,\beta)\right) \in \mathbb{R}^{m\times m}$

$(i,j)^{\underline{\text{th}}}$ element: $\Upsilon_{\text{iid},ij}(\rho,\beta) = \mathcal{G}_{n-m+i+j-1,1}\left(\frac{\gamma}{\beta(1+\rho)}, 1, -N_{\text{c}}\rho + 1\right)$

$\mathbf{\Upsilon}^{(\beta)}_{\text{iid}}(\rho,\beta) \triangleq \frac{\partial}{\partial\beta}\mathbf{\Upsilon}_{\text{iid}}(\rho,\beta) = \left(\Upsilon^{(\beta)}_{\text{iid},ij}(\rho,\beta)\right) \in \mathbb{R}^{m\times m}$

$(i,j)^{\underline{\text{th}}}$ element: $\Upsilon^{(\beta)}_{\text{iid},ij}(\rho,\beta) = \frac{N_{\text{c}}\rho\gamma}{\beta^2(1+\rho)}\,\mathcal{G}_{n-m+i+j,1}\left(\frac{\gamma}{\beta(1+\rho)}, 1, -N_{\text{c}}\rho\right)$

$\mathbf{\Upsilon}^{(\rho)}_{\text{iid}}(\rho,\beta) \triangleq \frac{\partial}{\partial\rho}\mathbf{\Upsilon}_{\text{iid}}(\rho,\beta) = \left(\Upsilon^{(\rho)}_{\text{iid},ij}(\rho,\beta)\right) \in \mathbb{R}^{m\times m}$

$(i,j)^{\underline{\text{th}}}$ element: $\Upsilon^{(\rho)}_{\text{iid},ij}(\rho,\beta) = \frac{\beta}{1+\rho}\Upsilon^{(\beta)}_{\text{iid},ij}(\rho,\beta) - N_{\text{c}}\,\mathcal{G}_{n-m+i+j-1,2}\left(\frac{\gamma}{\beta(1+\rho)}, 1, -N_{\text{c}}\rho + 1\right)$

TABLE II

REQUIRED CODEWORD LENGTH L AS A FUNCTION OF SNR γ FOR I.I.D. MIMO CHANNELS ($\mathbf{\Phi}_{\mathrm{T}} = \boldsymbol{I}_{n_{\mathrm{T}}}, \mathbf{\Phi}_{\mathrm{R}} = \boldsymbol{I}_{n_{\mathrm{R}}}$) AT A RATE 8.0 BITS/SYMBOL WITH $P_e \leq 10^{-6}$ FOR DIFFERENT NUMBERS OF ANTENNAS AND $N_{\mathrm{C}} = 5$

SNR (dB)	Codeword length L		
	$n_{\mathrm{T}} = n_{\mathrm{R}} = 2$	$n_{\mathrm{T}} = n_{\mathrm{R}} = 3$	$n_{\mathrm{T}} = n_{\mathrm{R}} = 4$
8	-	-	510
10	-	10865	75
12	-	210	30
14	-	65	15
16	1070	30	10
18	205	20	5
20	90	15	5

Note: The ergodic capacity $\langle C \rangle$ of 8.0 bits/symbol is attained at $\gamma = 14.55$ dB for $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$; $\gamma = 9.68$ dB for $n_{\mathrm{T}} = n_{\mathrm{R}} = 3$; and $\gamma = 6.79$ dB for $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$, respectively.

TABLE III

REQUIRED CODEWORD LENGTH L AS A FUNCTION OF CHANNEL COHERENCE TIME N_{C} FOR I.I.D. AND EXPONENTIALLY CORRELATED MIMO CHANNELS AT A RATE 8.0 BITS/SYMBOL FOR $P_e \leq 10^{-6}$, $n_{\mathrm{T}} = n_{\mathrm{R}} = 3$, AND $\gamma = 15$ dB

Coherence time N_{c}	Codeword length L	
	i.i.d.	$\zeta_{\mathrm{T}} = 0.5$, $\zeta_{\mathrm{R}} = 0.7$
1	18	53
2	24	68
3	30	84
4	36	100
5	45	115
6	48	126
7	56	140
8	64	160
9	72	171
10	80	190

TABLE IV

REQUIRED CODEWORD LENGTH L AS A FUNCTION OF CORRELATION COEFFICIENT ζ FOR EXPONENTIALLY CORRELATED MIMO CHANNELS WITH $\zeta_{\mathrm{T}} = \zeta_{\mathrm{R}} = \zeta$ AT A RATE 8.0 BITS/SYMBOL FOR $P_e \leq 10^{-6}$, $n_{\mathrm{T}} = n_{\mathrm{R}}$ =3, $N_{\mathrm{c}} = 5$, AND $\gamma = 15$ dB

Correlation coefficient ζ	Codeword length L
0.0	45
0.1	45
0.2	45
0.3	50
0.4	60
0.5	75
0.6	105
0.7	200
0.8	1015
0.9	-

Note: For $\zeta_{\mathrm{T}} = \zeta_{\mathrm{R}} = 0.9$, the ergodic capacity $\langle C \rangle$ is 7.36 bits/symbol at $\gamma = 15$ dB.

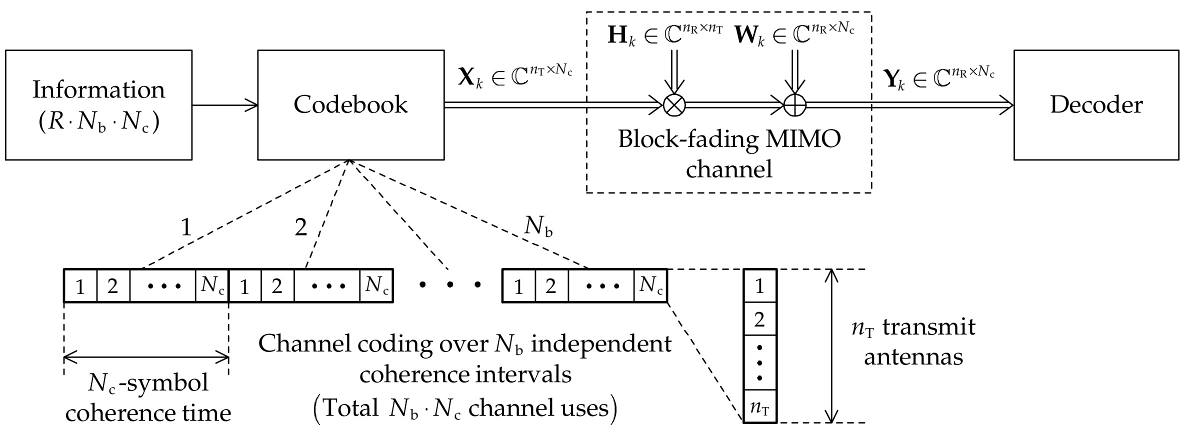

Fig. 1. A wireless communication link with n_T transmit and n_R receive antennas to communicate at a rate R over N_b independent N_c-symbol coherence intervals.

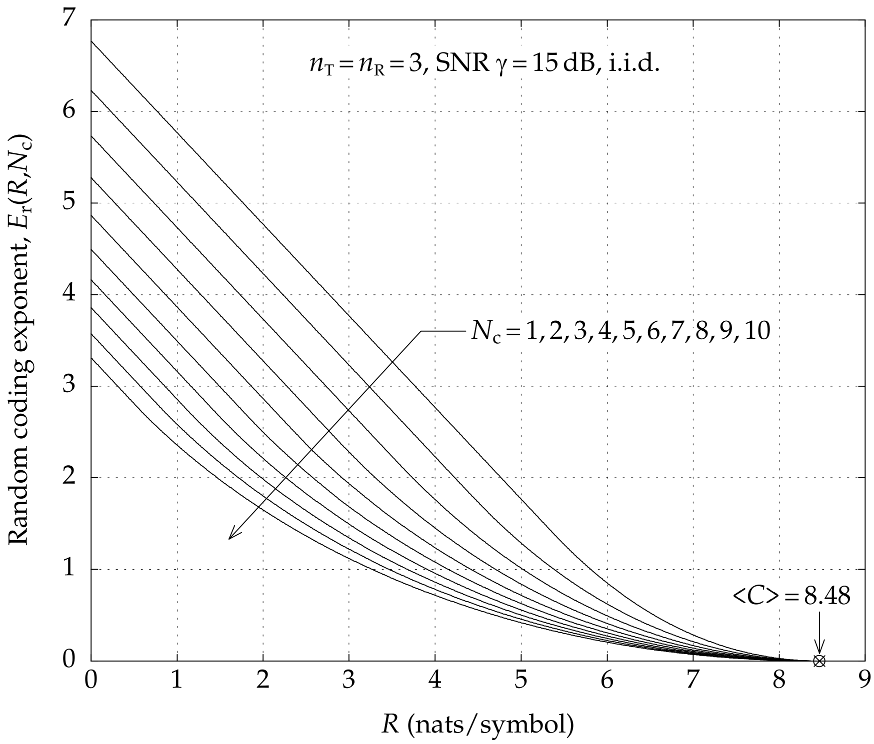

Fig. 2. Random coding exponent for i.i.d. MIMO channels ($\zeta_T = 0$, $\zeta_R = 0$) when $N_c = 1$, 2, 3, 4, 5, 6, 7, 8, 9, and 10. $n_T = n_R = 3$ and $\gamma = 15$ dB.

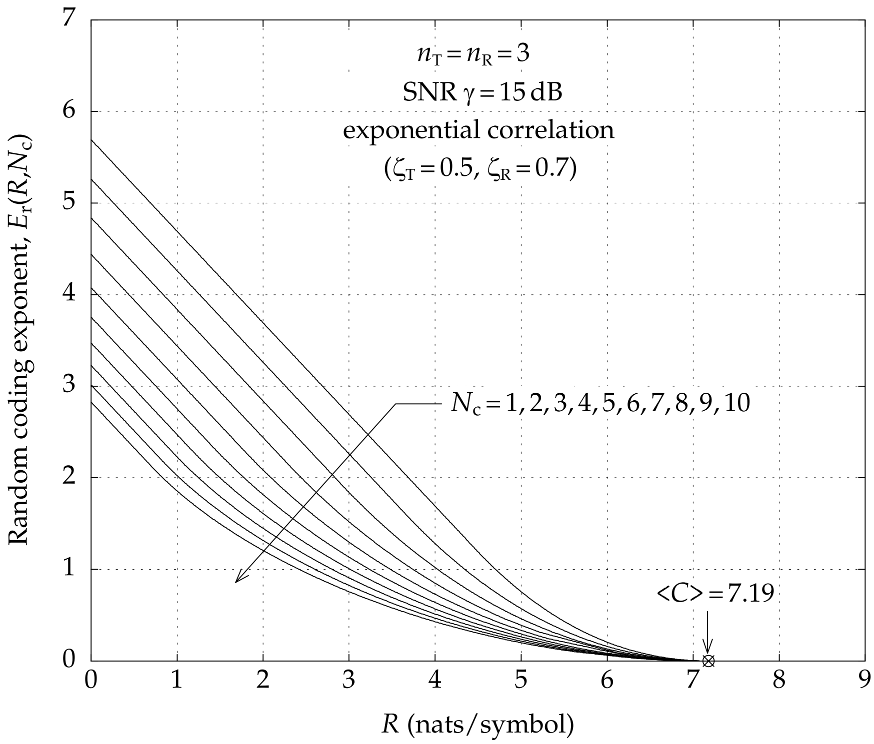

Fig. 3. Random coding exponent for exponentially correlated MIMO channels with $\zeta_\mathrm{T} = 0.5$ and $\zeta_\mathrm{R} = 0.7$ when $N_\mathrm{c} = 1$, 2, 3, 4, 5, 6, 7, 8, 9, and 10. $n_\mathrm{T} = n_\mathrm{R} = 3$ and $\gamma = 15$ dB.

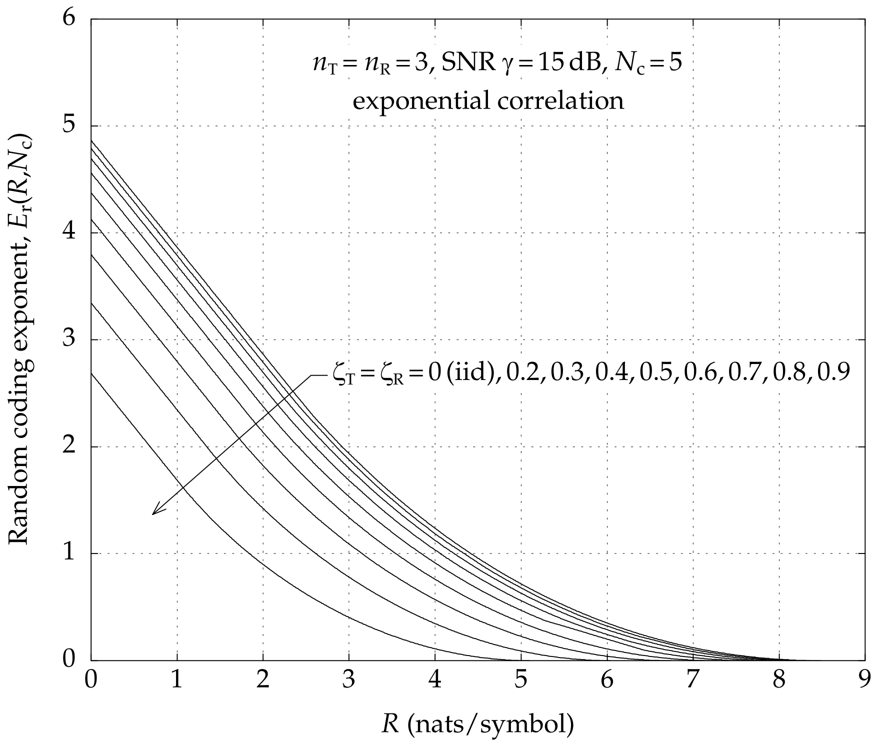

Fig. 4. Random coding exponent for exponentially correlated MIMO channels when $\zeta_{\mathrm{T}} = \zeta_{\mathrm{R}} = 0$ (i.i.d.), 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. $n_{\mathrm{T}} = n_{\mathrm{R}} = 3$ and $\gamma = 15$ dB.

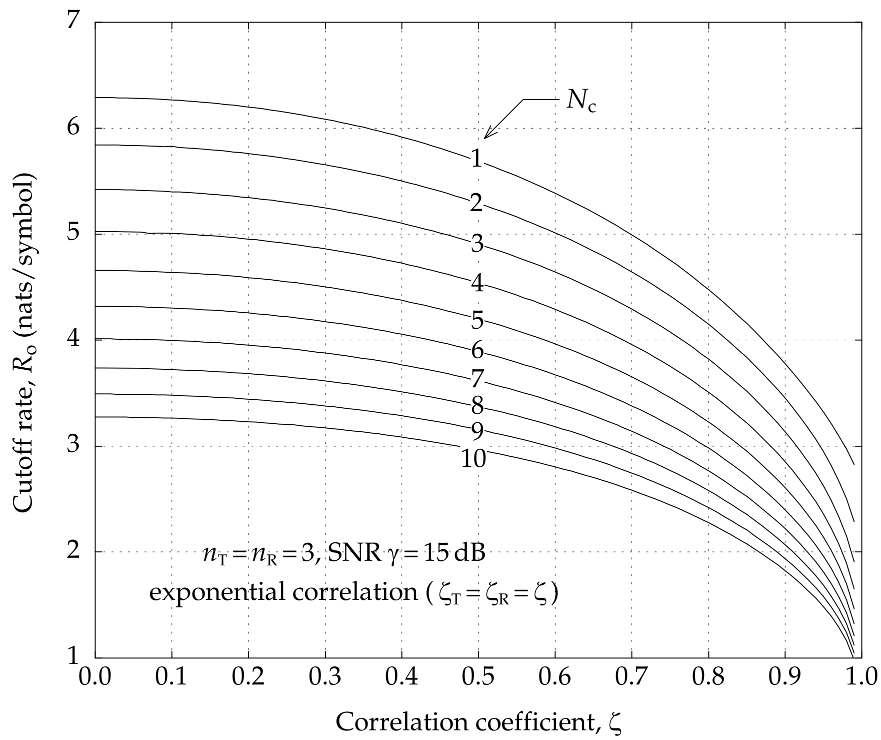

Fig. 5. Cutoff rate R_0 in nats/symbol as a function of a correlation coefficient ζ for for exponentially correlated MIMO channels with $\zeta_T = \zeta_R = \zeta$ when $N_c = 1, 2, 3, 4, 5, 6, 7, 8, 9,$ and 10. $n_T = n_R = 3$ and $\gamma = 15$ dB.